Public engagement and policy-making in the era of big data

Rempel, Emily

Award date:
2018

Awarding institution:
University of Bath

Public engagement and policy-making in the era of big data

Rempel, Emily

Award date:
2018

Awarding institution:
University of Bath

Public engagement and policy-making in the era of big data

Emily Suzanne Rempel

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Psychology and the Institute for Policy Research

July 2018

RESTRICTIONS ON USE

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation with effect from _____*(date)*

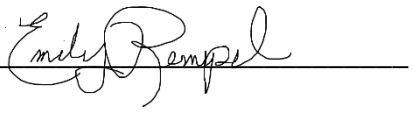
Signed on behalf of the Faculty of Humanities and Social Sciences

_____

LICENSE

DECLARATION OF MATERIAL FROM A PREVIOUSLY SUBMITTED THESIS/PORTFOLIO

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Candidate's signature 

DECLARATION OF AUTHORSHIP

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of the two articles presented in Appendix C: Additional Published Works where 20% of the work on the paper titled "Preparing the prescription: A review of the aim and measurement of social referral programmes" and 70% of the work on the paper titled "Realising the benefits of integrated data for local policy-making" was carried out by other researchers. I designed, carried out, and wrote up the former article, while I edited and carried out the data analysis in the latter. These articles are presented summarised alongside other external work in Chapter 9.

Candidate's signature 

TABLE OF CONTENTS

Page

Copyright Notice 1
Restrictions on Use 2
License 2
Declaration of material from a previously submitted thesis/portfolio 3
Declaration of authorship 4
Table of Contents 5
Dedication 12
Acknowledgements 13
List of Tables 14
List of Figures 15
List of Abbreviations 16
Abstract 17
CHAPTER 1 18
Introduction 18
1.1 Introduction 19
1.2 Research question and objectives 20
1.3 Academic disciplines and a brief introduction to the literature 20
1.3.1 Academic disciplines 20
1.3.2 Publics and public engagement 21
1.3.3 New data technologies 23
1.3.4 Government and data 24
1.3.5 Civic datafication 25
1.4 Overview of thesis chapters 25
1.5 Motives and arguments 29
CHAPTER 2 31
Methodology 31
2.1 Introduction 32
2.2 Epistemology and Ontology 32
2.3 Theoretical Constructs 34
2.3.1 Feminist theory 35

2.3.2 Decentred governance and theories of policy-making 36
2.4 Research Methods 39
2.4.1 Ethical considerations of research methods 41
2.5 Conclusion 43
Linking Text 44
CHAPTER 3 45
How should we do engagement? A literature review of public engagement for government data science 45
Chapter Declaration 46
3.1 Introduction 47
3.2. Background 49
3.2.1 Definitions of data science 49
3.2.2 Governments, citizens, and the future of data science 50
3.2.3 Early evidence of public disquiet around government use of data 51
3.3 Methodology 52
3.4 A review of public engagement with new technology 55
3.4.1 How to define the 'public' 56
3.4.2 A knowledge deficit and other early science-citizen interactions 57
3.4.3 Trust and trustworthiness in public engagement 58
3.4.4 Linking the aims and methodologies of public engagement 61
3.4.5 Imaginaries of the future for data science engagement 63
3.5 Propositions for government-driven data science 64
3.5.1 Proposition One: Consider the varied and many 'publics' who may be engaged in government data science 65
3.5.2 Proposition Two: Do not assume providing publics with information on data science initiatives will lead to public acceptance 66
3.5.3 Proposition Three: Determine the contingencies of trust for government data science and public engagement through trustworthy practice 67
3.5.4 Proposition Four: Design public engagements that incorporate robust, critical, and ongoing deliberation of data science 69
3.5.5 Proposition Five: Ensure holistic public participation that moves beyond privacy and consent 70
3.6 Conclusion 71
3.6.1 Strengths, limitations, and academic contribution 71

3.6.2 Final thoughts 72
References 74
Additional Data 81
Linking Text 84
CHAPTER 4 86
The 'hidden curriculum' of public engagement with data science 86
Chapter Declaration 87
4.1 Introduction 88
4.2 Methods and context 90
4.2.1 The ethnographic site 90
4.2.2 Methodology 90
4.3 Results and discussion 91
4.3.1 Lesson 1: Defining the public/Imaginations are made reality 92
4.3.2 Lesson 2: Setting the purpose of the Dialogue/Developing explicit and meta-objectives 93
4.3.3 Lesson 3: Engaging engagement/The meaning of engagement is contested 94
4.3.4 Lesson 4: Negotiating stakeholder interests 95
4.3.5 Other lessons and practices of interest 95
4.3.6 Reflecting on the author's role in the engagement 96
4.6.7 Critical reflection 97
4.5 Conclusion 98
References 100
Linking Text 103
CHAPTER 5 105
Story completion as an exploratory method into the rhetoric around data use in local UK government 105
Chapter Declaration 106
5.1 Introduction 107
5.1.1 Data in (local) government 108
5.2 Story completion as a methodology 109
5.2.1 What is story completion? 109
5.2.2 Designing a story completion study 111
5.2.3 The boundaries and benefits of story completion for civic datafication 112

5.3 Pilot study 114
5.3.1 Pilot study methodology 114
5.3.2 Pilot study results 115
5.4 Discussion of results and story completion 119
5.4.1 How do the stories compare to other literature on data sharing in local government 119
5.4.2 The viability of story completion in uncovering data rhetoric 121
5.4.3 Limitations and recommendations for story completion 122
5.5 Conclusion 123
References 124
Linking Text 126
CHAPTER 6 127
"So it's everything, data is everything": A qualitative study of the barriers and facilitators to data science in local UK government 127
Chapter Declaration 128
6.1 Introduction 129
6.1.1 Local government and data use 130
6.2 Methods 132
6.2.1 Design 132
6.2.2 Sample 133
6.2.3 Procedure and analysis 133
6.3 Results and discussion 134
6.4 The landscape of data projects in local government 135
6.4.1 Data as administration 135
6.4.2 Data as a narrative 136
6.4.3 Data as conferring objectivity 137
6.5 Facilitators and barriers of data projects in local government 138
6.5.1 The ebbs and flows of data projects challenge timescales 138
6.5.2 Challenges in the scope and kind of data available 140
6.5.3 Unclear guidelines encourage hesitance towards data sharing 141
6.5.4 Challenges in resources: skills and software 142
6.5.5 Relationships are key 144
6.5.6 The value of 'playing' with data 145

6.6 Conclusion 146
6.6.1 Final thoughts on civic datafication 146
6.6.2 Strengths, limitations, and critical reflection 147
6.6.3 Conclusion 148
References 150
Linking Text 152
CHAPTER 7 154
Contrasting publics' views of public engagement on local government data use: A qualitative study 154
Chapter Declaration 155
7.1 Introduction 156
7.1.1 Public engagement and data use 157
7.2 Methodology 159
7.2.1 Design 159
7.2.2 Sample 159
7.2.3 Procedure 160
7.2.4 Analysis 160
7.3 Results and discussion 161
7.3.1 Non-technical workshop themes 162
7.3.3 Technical workshop themes 166
7.3.4 Government workshop themes 170
7.4 Conclusion 179
References 180
Linking Text 183
CHAPTER 8 186
Exploring public preferences for engagement with UK government data usage: A quantitative analysis 186
Chapter Declaration 187
8.1 Introduction 188
8.1.1 Government and data technologies in the UK 189
8.1.2 The current state of public engagement on government data usage in the UK 190
8.1.3 Aims and objectives 192
8.2 Measures and methods 193

8.2.1 Overview 193
8.2.2 Main variables and measures 193
8.2.3 Engagement preferences and intensity 193
8.2.4 Psychological data ownership 195
8.2.5 Types of data 196
8.2.6 Other variables and measures 196
8.2.8 Parametric and non-parametric tests of difference 197
8.2.9 Mediation analysis 197
8.3 Results 198
8.3.1 Population description and randomisation validity 198
8.3.2 Categorical comparisons for public engagement preferences 199
8.3.3 Mediation of public engagement preferences and type of data 203
8.4 Discussion 204
8.4.1 Summary of results and comparisons to previous research 204
8.4.2 Reflecting on the contingencies of public engagement preferences 205
8.4.3 Strengths and limitations 207
8.5 Conclusion 207
Appendix 1: Additional variables and analysis 209
Description of additional survey variables 209
Perceptions of personal and public efficacy and trust in relation to government 209
Political behaviours 210
Perceptions of citizenship and interest in politics 210
Digital skills 211
Preliminary analysis 211
Preliminary results modelling public engagement preferences 212
Preliminary discussion on nuances of public engagement 212
References 217
CHAPTER 9 223
Additional Research 223
9.1 Introduction 224
9.2 Findings from external research 224
9.3 Conclusion 227

CHAPTER 10 228
Discussion 228
10.1 Introduction 229
10.2 A synthesis and summary of results 229
10.2.1 Objective One 229
10.2.2 Objective Two 236
10.2.3 Objective Three 241
10.2.4 Final key contributions to knowledge 244
10.3 Critical and theoretical reflections 247
10.4 Overall strengths and limitations 250
10.5 Future directions of research 252
CHAPTER 11 254
Conclusion 254
11.1 Personal reflections and final thoughts on reflexivity 255
11.2 Conclusion 257
References 259
Appendix A: Research Tools 264
Chapter 5 Story completion exercise 265
Chapter 6 Interview schedule 266
Chapter 7 Workshop schedule 268
Chapter 7 Workshop feedback form 269
Chapter 8 Survey instruments 271
Appendix B: Ethics Approval 283
Chapter 4 Research ethics approval and memorandum of understanding 284
Chapter 5 And 6 Research ethics approval, information sheets, and consent forms 287
Chapter 7 Research ethics approval and consent form 291
Chapter 8 Research ethics approval, information sheets, and consent form 294
Appendix C: Additional Published Works 298
Article: Preparing the prescription: A review of the aim and measurement of social referral programmes. 299
Article: Realising the benefits of integrated data for local policy-making 309

DEDICATION

I dedicate this thesis to my parents, Maria Kathleen Pike Rempel and Garry Edward Rempel. They have taught me to hold a strong sense of justice, to cherish inquisitiveness of spirit, to put empathy for others above all, and to value creativity and imagination.

Anything that is good in me is because of them.

ACKNOWLEDGEMENTS

First, thank you to my supervisors Professor Julie Barnett and Doctor Hannah Durrant. They are astounding, amazing, and brilliant people and academics. They have challenged me, made me work hard, but also made me laugh. They are the best possible supervisors anyone could ask for.

Second, I would like to acknowledge the immense support of my family and friends. In particular my sister Pamela Rempel and my best friend Karen Mok who have both listened to many long rants about my PhD but also which Disney Park we should visit next. I'd also like to thank my friends who are family, most particularly Louella Makortoff, Monique McRae, and Graham McRae. Also my fellow thesis survivors and friends Richard Hamshaw and Gigi Hoi. And finally to all of my family and loved ones both here and abroad.

And third, I'd like to thank all of my collaborators, co-workers, mentors, and participants for their support and guidance in finishing my thesis. In alphabetical order I would like to acknowledge Jack Close, Cat Drew, James Harle, Professor Nick Pearce, Jon Poole, Joe Prince, Caroline Ransford, Amy Thompson, Claudia Torres-Cervantes, Marsha Wood and many more. As well I'd like to thank my office mates and other students for their laughter and sometimes controversial discussion about the best member of the Queer Eye cast.

Finally, a quick recognition of some organisations and people who may not know me personally but have all helped me focus on things that are not my PhD. Thank you to the Theatre Royal Bath's Theatre School, the Walt Disney Company's many theme parks and films, Nick Mason and James Clement from The Weekly Planet, and the Odeon Cinema Bath.

LIST OF TABLES

Table 2.1: Methodological and analytic techniques of thesis chapters

Table 3.1: Search terms and databases

Table 3.2: Initial and final thematic analysis of public engagement literature results

Table 3.3: Literature review lessons and propositions for government data science in the UK

Table 6.1: Thematic analysis results of 11 qualitative semi-structured interviews

Table 7.1: Thematic analysis results

Table 8.1: Public engagement preference scale

Table 8.2: Dimensions of type of data used in survey randomisation

Table 8.3: Chi-squared, Kruskal-Wallis, and ANOVA testing of any difference in demographic distribution by type of data

Table 8.4: Chi-squared, Kruskal-Wallis, and ANOVA testing of any difference in distribution of public engagement preferences and psychological ownership by type of data

Table 8.5: Mediation difference testing using simple linear regression

Table 9.1: External research projects on the topic of government data use

Table 10.1: Main findings from thesis

LIST OF FIGURES

Box 5.1: Example Story Completion Stems

Figure 8.1: Direct acyclic graph for the proposed relationship between engagement preferences and intensity, perceived ownership of data and type of data.

LIST OF ABBREVIATIONS

EU	=	European Union
FOI	=	Freedom of Information
GDPR	=	General Data Protection Regulations
GDS	=	Government Digital Service
IG	=	Information Governance
IQR	=	Interquartile Range
LGA	=	Local Government Association
MP	=	Member of Parliament
ODI	=	Open Data Institute
ONS	=	Office for National Statistics
P	=	P-value
PEST	=	Public Engagement with Science and Technology
PUST	=	Public Understanding of Science and Technology
R^2	=	Coefficient of Determination/R-Squared value
REF	=	Reference Value (see: Regression Tables)
SD	=	Standard Deviation
UK	=	United Kingdom
USA	=	United States of America

ABSTRACT

Data design, collection, and use are increasingly part of public life in the United Kingdom. Policy-makers in particular are engaging with these kinds of data technologies to increase efficiencies and cut costs in public service provision. Data technologies are subsequently increasingly important in mediating the processes and relationships of communities, citizens, and government. This thesis aims to examine how publics and policy-makers can engage with emerging data technologies in the context of local and national government in the UK. The objectives are as follows. First, to summarise and evaluate the current research and state of government public engagement with data technologies. Second, to scrutinise the current use of data technologies in government through qualitative semi-structured interviews and other qualitative methods. And third, to examine the potential for new forms of public data engagement in government through qualitative workshops and analysis of public views on government and data technologies. This thesis uses a mixed methods approach, which is primarily qualitative, to explore these objectives including a narrative literature review, ethnography, semi-structured interviews, story completion exercises, focus groups and statistical modelling through difference mediation analysis. Key findings include, first, that current public engagement practice on government data use is often limited in topic and form. For example, small-group discussions on public views of privacy that have no clear line to impact policy change. Second, that data use in government is subject to political motivation that generally takes the form of narrative testing. And third that public engagement in the era of big data in government must focus both on means and ends. Government must move beyond a model of seeing 'the public' as a roadblock on the way to civic datafication and instead as an integral part of debating the components of ethical and socially relevant data practice.

CHAPTER 1

Introduction

1.1 INTRODUCTION

Governments around the world have embraced data technologies as a mechanism to find a solution to nearly every misfortune imaginable: from poor health and poverty to crime and climate change. Data technologies are imagined as our salvation. However, they have also been foreshadowed as our downfall. In fiction, imaginations of the future are often dystopian in nature with governments that track and monitor their citizens on the most finite of details (Dalrymple, 2001; Gill, 2017; Lepore, 2017). The use of data is imagined as a tool of government control, generally to the peril of citizens (Spielberg, 2002). Despite these terrifying projections, they are not reality. Government does not yet have the capacity to predict future murders or track our every thought, nor on a fundamental level have the rules been set on what government *should* do with these kinds of data technologies (Malomo & Sena, 2017).

Data production and analysis is, however, at an all-time high. New kinds of data from internet sources like social media websites are rapidly expanding. Five hundred million tweets are sent each day across the globe (Internetlivestats.com, 2018). As of March 2018, the UK government recorded 3.31 billion citizen and business transactions annually, representing on average 5042 transactions per person per year (Gov.UK, 2018). Everything from Facebook to bin collection to coupon allocation is built upon these data. There is an unprecedented phenomenon in calls for 'smart cities' and 'business analytics', encouraging both public and private enterprise to use data as the building blocks of their organisations (Keller, Koonin, & Shipp, 2012). Entirely new fields of academia have been established including critical data studies, data science, and big data analytics. And yet we have not addressed fundamental questions about what this means for government, what this means for communities, and what this means for us as individuals. And in particular how to ensure the future of data use in government is for the benefit of citizens rather than a dystopian premonition come true.

This thesis aims to explore, and hopefully answer, some of these questions. The introduction is organised as follows. First, the overall research question and objectives are described. Second, the three academic disciplines drawn upon are described and previous literature is briefly outlined. Third, an overview of subsequent chapters is presented. And fourth, summative arguments on the motives behind this research are discussed.

1.2 Research question and objectives

The aim of this thesis is to examine how publics and policy-makers can engage with emerging data technologies in the context of local and national government in the UK. This work seeks to address the following objectives:

1. To summarise and evaluate the current research and state of government public engagement with data technologies.
2. To scrutinise the current use of data technologies in government through qualitative semi-structured interviews and other qualitative methods.
3. To examine the potential for new forms of public data engagement in government through qualitative workshops and analysis of public views on government and data technologies.

1.3 Academic disciplines and a brief introduction to the literature

1.3.1 Academic disciplines

This thesis represents a programme of interdisciplinary work. It draws on theories and traditional academic disciplines related to political science, psychology, and public health. It also utilises specialisations of these disciplines including public engagement, critical data studies, and science communication. To a minor degree, theories from media studies, science and technology studies, education research, and sociology are also drawn upon to frame the research perspectives and methods. This work, therefore, does not build upon an established area of research but rather synthesises a broad range of social sciences disciplines to propose a new area of specialisation named civic datafication

studies. This term is novel and is discussed in detail in Chapters Five and Six in this thesis. Civic datafication includes three main dimensions: publics, new data technologies, and government. These three dimensions are briefly discussed in relation to civic datafication in the next section. A further detailed review of the literature is presented in Chapter 3 as an empirical paper.

1.3.2 Publics and public engagement

Publics are varyingly defined as both fluid and finite groups of individuals. On the one hand, Newman (2011) argues that publics are constructed and dynamic in the political context where leaders call upon and summon publics in relation to specific issues. Renn (2008) contrastingly describes publics within four finite pre-existing categories as stakeholders, affected publics, observing publics, and the general public. In the context of public engagement with data technologies, publics can also be self-identified or defined. For example, civic hacker groups who self-identify themselves as public data technology activists. Data can be used to create publics, for example Ginnis et al.'s (2016) algorithmic determination of 'data adopters' versus 'data pragmatists'. This thesis, from an epistemological perspective, considers publics to be fluid and dynamic groups that are defined and constructed through various processes including political summoning, behaviour targeting, self-identification and more. Publics are therefore groups of individuals defined by their shared characteristics and identifications. They are fluid because these characteristics and identifications can change over time. Individuals can belong to many or few publics and the size of publics can vary by the characteristics used to classify them. For example, residents who live in the South West of England and are civically active around climate change may be a relatively large self-identified public whereas smartphone users who attend the University of Bath would be a relatively small and more strictly defined public. Of specific interest to this thesis are government publics, e.g. policy-making and civil servant publics, that are involved in using data and designing data-related projects, non-technical local publics that data are collected upon, and technical publics that specialise in designing or using data

with government or in private industry. These publics are of particular interest because of their close relation to the issues related to government data use. These publics often hold contrasting and varying levels of power in relation to these characteristics and the intersections of these characteristics (Davis, 2008). For example, publics who collect data like governments and technical elites have more power over the control of that data than publics who data are collected upon. This thesis considers publics to be many and varyingly defined to avoid definitions like the 'general public' that often obscure the variations between individuals as well as the inherent power related to different classifications. In sum, this thesis defines publics as fluid rather than fixed to highlight the dynamic nature of publics, the varying levels of power between publics, and to avoid over-simplifications like the 'general' public.

Public engagement with science and technology is any activity between the publics who create and govern technology and the publics who are not related to the development and governance of said technology. Highlighted in this definition, is the notion that the publics who govern technologies are not the 'experts' or 'stakeholders' as described by authors like Renn (2008). These latter terms consist of latent assumptions of assumed importance and power for experts as they are set above and outside of the 'general public'. Thus the conception of public engagement with science and technology as described in this thesis reflects the definition of publics as fluid and dynamic. This thesis draws from public engagement with science and technology literature rather than participative democratic theory. A key question to consider in defining the boundaries of this work is the differences between public engagement and the democratic function and activities of government. Indeed how participative democratic activities like community consultation can be distinguished from public engagement. This thesis considers public engagement with science and technology to be a subcategory of democratic activity which specifically focuses on the practical mechanisms of including publics in the development and use of new technologies. Public engagement, as a practice, draws to some degree on theories of democracy like

decentred governance (see: Braun and Konninger (2017)). However, public engagement literature focuses more on the social aspects of technical development and society in general rather than political theory (Stilgoe, Lock, & Wilsdon, 2014). As this thesis is centrally located on the relationship between data technologies and publics in the context of government, this thesis draws on public engagement literature that problematises notions of technology and society rather than notions of democratic reform seen in the political science literature.

Engagement activities can be grouped into three broad categories defined by their mode of interaction (Rowe & Frewer, 2005). First, communication which is one-way communication from developer to publics. Second, consultation defined as one-way communication from publics to developers, with no feedback from developers to publics on the result of the consultation. And finally, collaboration which is two-way communication between developers and publics. This categorisation does not include activities that may be driven by publics themselves, for example protest and citizen science activities like civic hacking. Public engagement can also be defined by the aim of engagement. Fiorino (1990) describes these aims as instrumental, substantive, or normative. Instrumental engagements include publics as a requirement of technology development to ensure the success of that technology, substantive engagements include publics to improve a technology, and normative engagements include publics as they are due a role in technologies that may impact them. Public engagement therefore covers a broad range of activities that aim to include non-technical publics in some aspect of science or technology.

1.3.3 New data technologies

Everything from how we communicate socially to how government runs services is being reimagined and redeveloped in light of data and related data technologies. New data technologies relate both to new forms of data as well as new applications of data. This includes technologies like big data, open data, and data science. From a technical viewpoint open data is data released in the public

domain (Hand, 2013), while big data refers to massive datasets that are automatically and routinely collected (McNeely & Hamh, 2014; Ward & Barker, 2013). Data science refers to the combination and application both of traditional 'small' data as well as these new forms of data. Data technology applications also include advances in analysis and algorithms ranging from the highly complex like blockchain technologies to the relatively simple like predictive texting (Piscopo, Siebes, & Hardman, 2017). However these technical definitions tend to ignore the social aspects introduced by data technologies where "Big Data [is viewed] as a cluster or assemblage of data-related ideas, resources, and practices" (Markus & Topi, 2015). This social perspective better allows for exploration on how data is incorporated both in government and day-to-day life and accordingly will be used in this thesis.

1.3.4 Government and data

The UK government consists of a multitude of public institutions, represented by central, local, and devolved governments, responsible for the regulation and organisation of the state (Parliament.UK, 2018). Of particular interest to this thesis is the contribution of data technologies to the ideation and formation of policy choices but also the administrative function of the national government's civil service and local government. This thesis focuses on government public engagement rather than industry public engagement with data technologies due to the precarious position of government being both a developer and regulator of data use. The UK government is highly invested in driving towards a model of 'digital by default' (Cabinet Office & Government Digital Service, 2017). Government, therefore, are developing these new data technologies while also establishing the rules and regulations for their use. As governments also have a mandate to be responsible to the citizens they govern, this thesis focuses specifically on public and policy engagement with new data technologies. Further reflection on theories related to democracy, governments, and data use are presented in Chapter 2: Methodology.

1.3.5 Civic datafication

Finally, this thesis seeks to describe the overlap of and interplay between publics, governments, and data technologies through the introduction of a novel term. As described above several authors have attempted to define new data technologies both for their technical and social aspects (Markus & Topi, 2015). Rather than re-defining these pre-existing terms that are often fluid and contested, this thesis introduces a new term that seeks to describe this public-government-data relationship, namely *civic datafication*. This is conceptualised as the ways in which data and data technologies mediate and define the interactions between publics and governments. While datafication is a fairly commonly used term in critical data studies to describe the increasing use of data and digital technologies (Data Power, 2017; Techopedia Inc., 2018) this thesis chooses to integrate the word civic to represent the ways in which publics and government are centrally located in the consideration of data technologies in this thesis. Therefore *civic datafication* is defined as the dramatic increase in the importance of data technologies to mediate the processes and relationships of communities, citizens, and government. The concept of civic datafication is explored more fully in Chapter 5.

1.4 OVERVIEW OF THESIS CHAPTERS

This thesis is presented in the alternative paper format with ten chapters, of which seven are empirical research studies. Each chapter, excluding the methods and discussion but including the literature review, are formatted as journal articles. As this thesis is in the alternative paper format, a certain amount of replication within the introduction to each chapter as well as in the individual discussion sections is possible. Chapter 3 has been accepted for publication while Chapter 4 has been submitted for publication. Chapters 5 through 8 are formatted as articles but not yet submitted for publication at the time of thesis completion. Each chapter consists of an individual research project and seeks to address one of the core objectives described in Section 1.2. Materials used in the studies, as well as descriptions of data sources, are presented in the Appendices and individual methods sections of each paper.

Chapter 2: Methodology

In Chapter 2 philosophical and theoretical clarity for the overall thesis is established. This first includes an overview of the epistemological and ontological positions of the research, namely a contextualist and pragmatic approach. This is followed by a reflection on the interdisciplinary nature of the thesis in relation to theories of feminism and governance. While each individual chapter outlines its methodology, this section's aim is to describe how these methodologies work together and discuss how they relate to the epistemological rationale of the thesis. A range of methods are presented, describing the mostly exploratory and qualitative approach, in relation to the objectives described in the introduction and the theoretical underpinnings of the work. The chapter closes by reflecting on the ethical considerations of the methodology.

Chapter 3: How should we do engagement? A literature review of public engagement for government data science

In Chapter 3 current literature on public engagement is synthesised and lessons for government on how to incorporate the history of public engagement into the future of public engagement with data technologies is proposed. The overall aim is to examine the potential for public engagement with government data science through a critical analysis of engagement and trust building in other new technologies. This chapter offers an in-depth review of the literature around data technologies, government, and public engagement. It subsequently critically reviews both the history of public engagement from governments around the world as well as past citizen engagements around data technologies. This chapter seeks to address Objective 1.

Chapter 4: The 'hidden curriculum' of public engagement with data science

In Chapter 4, an ethnographic study of a national UK government data science public engagement is presented. This study's aim is to examine an example of how public engagement with data science was operationalised. This chapter draws on public engagement literature and the theoretical concept of the

'hidden curriculum'. This perspective was chosen because the hidden curriculum is a useful tool for using ethnography to study the underlying mechanisms that influence social activities. Although it is traditionally employed in relation to education research, it was used here to critically evaluate the 'status quo' of public engagement. The theory was applied to a series of ethnographic notes collected from participatory observation of the Cabinet Office's Public Dialogue on Data Science Ethics conducted from late 2015 to mid-2016. This chapter also seeks to address Objective 1.

Chapter 5: Story completion as an exploratory method into the rhetoric around data use in local UK government

In Chapter 5 results from a study piloting the qualitative technique *story completion* to examine the contingencies of data projects in local UK government are presented. The objectives of this project are first to explore how local government workers imagine hypothetical data projects, second to explore how this may or may not reflect the contingencies of data projects in local government, and third to examine whether story completion is a viable method for understanding policy and data processes. As this study piloted a method new to the study of policy-making and narratives, it was deemed necessary to critically engage with the literature and evaluation of the method itself. Subsequently, this chapter also describes the story completion method in detail as a proposal for its relevance to studying social narratives around policy problems. This chapter seeks to address Objective 2.

Chapter 6: "So it's everything, data is everything": A qualitative study of the barriers and facilitators to data projects in local UK government

Chapter 6 is a sister study to Chapter 5 using the same participant group. In this chapter, semi-structured interviews around two core data projects in a local authority in the South West of England were conducted. The aim of this chapter is to explore the barriers and facilitators of data projects in local UK government using thematic analysis of qualitative interviews. This includes critical analysis of

the spaces for public engagement within local government data practice. This chapter also seeks to address Objective 2.

Chapter 7: Contrasting publics' views on government data use and public engagement in the UK: A qualitative study

In Chapter 7 the results from a series of focus groups and workshops with local government workers, technical publics, and non-technical publics are presented. The aim of this chapter is to compare and contrast differing public views on public engagement with local government data usage. A series of seven focus groups were conducted over four workshop sessions. The chapter brings together lessons from previous chapters to ask, qualitatively, what different publics desire for their engagement around government data use. This paper takes a critical, analytic approach to reflecting on the contingencies of government data practice as well as publics' own views on how future government public engagements around data technologies could run. This chapter seeks to address Objective 3.

Chapter 8: Exploring public preferences for engagement with UK government data usage: a quantitative analysis

In Chapter 8 a quantitative, empirical study is presented on the views of UK publics on the nuances and desires for public engagement around government data use. The aim of this study is to explore the relationship between different types of data, psychological ownership towards data, and public engagement preferences with data. 1100 UK residents completed an online survey on this topic. The first objective is to compare and contrast preferences for levels of engagement by type of data. The second objective is to use linear regression and difference mediation testing to examine if levels of ownership mediates any relationship between type of data and preferences for levels of engagement. The third objective is to develop a multivariate regression model for predictors of preferences for public engagement with government data use. This chapter seeks to provide recommendations for best practice in government data public engagement in

relation to the desires for engagement as expressed by publics themselves. This chapter also seeks to address Objective 3.

Chapter 9: Additional research

Chapter 9 presents results from two collaborative external research projects conducted by the author on the topic of government data use. As this research was not funded by the PhD nor was it solely undertaken by the author, it is presented and briefly summarised in this chapter. The two projects were both collaborative efforts between the Institute for Policy Research at the University of Bath and local governments to better use data in policy decision-making. Thus these projects provided an opportunity for the author to further explore the nature of data usage in local government, and particularly the political contingencies of its use. This chapter accordingly seeks to address Objective 2.

Chapter 10: Discussion

In Chapter 10 the findings from the preceding six empirical papers, alongside the additional research chapter, are synthesised and summarised. Lessons learned from these chapters are discussed and the overall contribution to knowledge in the social sciences is evaluated. The main aim and three objectives of the thesis and how these papers and ancillary projects have answered these objectives is discussed. The discussion ends with a reflection on strengths, limitations, and areas of future research.

1.5 MOTIVES AND ARGUMENTS

Beyond the objectives and aims highlighted in this introduction as well as subsequent empirical chapters, I have several core motives and arguments that led me to conduct this research. This subsection is written in the first person to highlight the reflexive nature of this section as well as to position myself within feminist scholarship as a researcher who is also a social advocate. I feel it is imperative to the quality and transparency of this research that I am clear why I chose to write my thesis on this topic. My previous research was in the area of epidemiology and health sciences using government big data analysis (Rempel,

2015; Rempel, Somers, Calvert, & McCandless, 2015). This work provided a window into the kinds of data collected by government and the guidelines for its use and application. It also demonstrated the scale and nature of data collected about individuals around the world, particularly vulnerable indigenous and offender populations who are subject to intense and disproportionate data surveillance (Data Power, 2017). I therefore sought to explore this topic in my doctoral research to help contribute to critically and qualitatively examining government use of data. My arguments are as follows:

1. First that data use and data technologies are developing faster than our social and critical evaluation of them.
2. Second that publics are often unaware both of the kinds of data analysis done on them, as well as the impacts of that analysis, by government, academia, and industry.
3. Third that individuals and publics have a fundamental right to know how their data is used and how that impacts both themselves and their communities.
4. Fourth that democratic governments, due to their mandate as both representing and serving publics, should be a leader in ethical and transparent data use that takes in to account public views and opinions.
5. And fifth that data technologies, for example surveillance using big data, should be understood as social objects that must be put under public and academic scrutiny.

In subsequent chapters I do not always explicitly address these arguments due to the nature of the paper format, but rather they are fundamental to the kinds of methods I chose as well as the conclusions I draw. In the discussion I will return to these five points to examine whether this research provided adequate evidence to support my overarching arguments.

CHAPTER 2

Methodology

2.1 INTRODUCTION

This chapter presents an overview of the thesis' research methodology in three sections. First, the epistemological and ontological rationale is presented. Second, key theoretical constructs are discussed. And third, the research methods used in this thesis are outlined and connected to the philosophical, theoretical and ethical rationale presented. As this thesis is presented in the alternative format, detailed description of the methods are included in each of the study chapters. All of the interview schedules, survey instruments, and other research tools are listed by study chapter in Appendix A: Research Tools. Copies of ethical approval emails are listed in Appendix B: Ethics Approval, all projects including human participants were approved by the University Of Bath Department Of Psychology's Research Ethics Board.

2.2 EPISTEMOLOGY AND ONTOLOGY

This thesis approaches social sciences research from a problem-based epistemology. This is what Mjøset (2009) names the contextualist approach and Onwuegbuzie and Leech (2005) the pragmatic approach. If epistemology is visualised as spanning two dichotomous poles from traditionally experimental and positivist, where research is only an activity of discovery, to social constructionist on the other, where research is only an activity of interpretation, then pragmatism and contextualism exist in the middle. From an ontological perspective, the nature of reality would be described as both interpretive and realist. In other words there are naturally-occurring phenomena that exist independent of human interpretation, e.g. molecules, cells, organs etc., but our understanding of these phenomena is deeply rooted in social interpretation and construction, e.g. gender, work, health, etc. Both contextualism and pragmatism highlight that research should see 'knowledge as problem driven' (Mjøset, 2009, p. 64). Put another way, choices made during the discovery and interpretation of social phenomena should be based on the context and characteristics of that phenomenon. Methodology is not ideally qualitative or quantitative, but rather certain research tools are more suited to certain research questions. Pragmatism

challenges the existing paradigm that quantitative research is only a tool for positivist or purist research and qualitative research is only a tool for constructionist research. One of the key benefits of this perspective is that 'pragmatic researchers utilize mixed methodologies within the same inquiry, they are able to delve further into a dataset to understand its meaning and to use one method to verify findings from the other method' (Onwuegbuzie & Leech, 2005, p. 384). Therefore the 'middle' epistemology emphasises employing a variety of techniques to explore, as fully as possible, a social problem. A problem-based epistemology is not an uncertain or compromised epistemological view but rather a challenging space in which the researcher must critically evaluate what methods are best suited to the research question at hand.

A strength of this perspective on epistemology is that it does not subscribe to the belief that context is divorced from good scientific practice. Good scientific practice is not about isolating social phenomena through experimental methods, e.g. lab-based studies, but about embracing and understanding the context that social phenomena exist within. Consideration of the context is integral to the middle-position, and particularly to building knowledge of process and outcome. In the middle perspective how we find something out, i.e. process, is as important as the outcome itself, and both are outcomes of the research results (Mjøset, 2009). Thus describing the context and the conclusions of the research are both essential to understanding the social problem at hand.

Contextualism in particular focuses on case-based methods, seeing theory-building as a dynamic process of building knowledge through repeated cases. This challenges the view that qualitative and case-based research is less valuable due to being less generalizable. Cases all form part of a larger paradigmatic view of social phenomena, and knowledge is built through comparing and contrasting cases (Mjøset, 2009). These epistemologies highlight the key role of social constructionism in research concepts like generalizability. Overarching conclusions about social phenomena or typologies 'should not be turned into

essential features of reality. Although they are empirically grounded, typologies are still constructions' (Mjøset, 2009, p. 63). Conclusions are not fixed in reality, but are constructions in and of themselves and should be reflective of the context within which they sit. New cases should always be examined to challenge the established view of social reality, regardless of method or size of study. The problem must be defined through and by context, thus the methodology chosen should be the one that best delves into the context rather than by what the researcher deems scientifically pure. The methodology should be appropriate to the problem first.

2.3 Theoretical Constructs

This thesis draws on two sets of theoretical constructs in conducting and analysing research: theory in practice and theory in approach. Theory in practice refers to theories and concepts used in the framing, design, analysis and discussion of the six thesis papers. Theory in approach refers to the underpinning theories used to construct the methods chosen as well as used in the synthesis of the overall findings. This distinction is drawn due to the structure of the thesis. Several theories including things like the hidden curriculum (Cotton, Winter, & Bailey, 2013) from education research and public-making (Newman, 2011) from governance and policy research were drawn on in specific papers as these were relevant to the discipline and topic of that paper. However they were not used in the overall discussion or understanding of the thesis as a whole. This fits within the pragmatic epistemology of the thesis as relevant theories were drawn on in respect to the individual problem of each chapter. These theories used in practice are described within their respective chapters, while the theories used in the approach are described below.

There are no established theories that encompass the entirety of the relationship between publics, government, and data technologies due both to its interdisciplinarity and novelty. And indeed there are no singular theories of public engagement. Public engagement instead draws on theories related to a

multitude of concepts in policy-making, governance, science communication, and epistemology. These concepts are described in detail in Chapter 3. As there is no appropriate single theory to frame and build this thesis upon, instead two theoretical domains that are most useful for exploring public engagement with government data use are presented: feminism and decentred governance. This thesis also, in line with contextualist and pragmatic thinking, focuses on using these theories to understand practice rather than test theories. This thesis is not conceptualised as a process of theory testing but rather as knowledge building through feminist activity.

2.3.1 Feminist theory

While this work is not a traditional piece of feminist scholarship in that it does not address problems of gender, feminism has much to add to the understanding of public engagement in government. Feminism has direct correlates to public engagement both in its recognition of the influence of power in politics as well as positioning the researcher as active within a social movement. Feminist scholarship draws upon intersectionality, interdisciplinarity and the interplay of activism and research (Ferguson, 2017). Davis (2008) defines intersectionality as 'the interaction between gender, race, and other categories of difference in individual lives, social practices, institutional arrangements, and cultural ideologies and the outcomes of these interactions in terms of power' (p. 68). Hence different aspects of social existence lead to different experiences of social life, and subsequently differing opinions and power in relation to social phenomena like technology development. Intersectionality represents a recognition of publics as pluralistic, which relates strongly to the view that public engagement enables the expression of different views of science and technology. Intersectionality also addresses the different levels of power within society and subsequently the challenges that different groups face in expressing their lived realities. Public engagement is also about recognising and challenging the lack of power that non-technical publics have in addressing the social aspects of

technology development. In sum, this research is based on a feminist perspective that public engagement is an activity of enabling and establishing voice in less powerful publics within the context of data use in government. This situates the research within feminist scholarship as feminism ultimately aims 'to trouble power relations, imagine better worlds, and work to achieve them' (Ferguson, 2017, p. 283). Research on public engagement draws on the same ideals.

In relation to situating the researcher in a social movement, Ferguson (2017) contends that '[f]eminist theory is a change-oriented scholarly practice; challenging oppression and working toward justice are not separate applications of a theory made elsewhere but constitutive elements of theory making' (p. 275). Public engagement is similarly, not merely an object of study but also a process to engage with. Studying public engagement becomes an activity of understanding and supporting the recognition of pluralistic views on data usage within and outside government. This positions the researcher as a proponent of change and not merely as the conduit through which to describe the social phenomenon of public engagement and policy-making in the era of big data. This is best reflected in the arguments described in Section 1.5 of the Introduction that position the research as part of a process of establishing the justification for increasingly engaging publics on the topic of data governance and data usage. Feminism's three main dimensions of intersectionality, interdisciplinarity, and activism are thus all key to understanding how to study and advance the field of public engagement.

2.3.2 Decentred governance and theories of policy-making

Also relevant to the role of public engagement in government are theories of state, democracy, and policy-making. This thesis does not focus on, nor do the chapters advance, any specific theory of the policy process or democracy. However, these concepts are nevertheless relevant to understanding how public engagement and government coincide and how governments use data both in administrative and policy-making functions. First, theories of policy-making are

discussed. The use of data in government, which this thesis proposes publics should be engaged in, is an activity related to policy-making and public administration. It is therefore important to reflect on how this thesis approaches the concept of policy-making in order to evaluate the potential for public involvement. Next, relevant theories related to democracy are discussed as public engagement in government is, itself, a form of democratic activity.

This thesis imagines policy-making as a process of competing and contested narrative-making where power and context influences the strength of that narrative to succeed. As is common in political science, this thesis problematises the concept of policy-making as a linear or circular process of agenda setting, policy selection, and policy reform (Cairney & Heikkila, 2014; Janssen & Helbig, 2016). A vast range of alternative theories of policy-making exist that challenge this normative linear pathway including things like punctuated equilibrium theory, narrative policy framework, and advocacy coalition framework (Cairney & Heikkila, 2014). All of these theories suggest that policy-making is a 'messy' process in which conflicting ideas and policies are brought forward, debated, and implemented but not always in that order.

This thesis most strongly aligns with a narrative-conflict view of policy-making as highlighted by authors like Deborah Stone (Stone, 1989). These theories highlight the role of actors and groups in making meaning out of policy problems by advancing narratives of causation. Narratives involve subjective and emotional decision-making, as Davidson (2017, p. 3) states "[s]ome of the most important policy debates, particularly in deeply divided, multicultural societies, involve decisions not only about means but also about ends. These are questions of what "ought to be"-questions of values, not facts." Stone's (1989) now classic description of causal theories similarly positions the actants within policy-making as attempting to shift the policy problem narrative into different intended and unintended consequences as well as unguided and purposeful actions. Policy problems that have intended consequences and purposeful actions have an

'intentional cause' and subsequently are deemed relevant to policy intervention. In Stone's (1989) own words, "[s]ince our cultural understanding of accidents defines them as events beyond human control, causal politics is centrally concerned with moving interpretations of a situation from the realm of accident to one of the three realms of control" (p. 274).

A few authors have begun to discuss how data and data technologies fit into a narrative-conflict process of policy-making. Kettl (2016) emphasises that the nonlinear nature of policy-making problematises the assumption that data is used simply as evidence to make the best policy choice. In fact they argue that good data analysis is useless without a good narrative provided. However, authors that focus more on practice than theory like Janssen and Helbig (2016) argue that data technologies have great potential to interrupt the status quo and revolutionise the traditional policy cycle as they see it. Scholars that study the use of data technologies in government are at the early phase of understanding how these technologies challenge or inform current political theories.

Finally this thesis incorporates theories of democracy and state, specifically focusing on decentred governance. Several authors highlight the waning representivity and success of modern advanced democracies (Achen & Bartels, 2016; Farrell, 2014). Similarly to theories of policy-making, several theories of democratic reform compete. These theories are particularly relevant to public engagement as they conceptualise of changing the relationship between citizens and decision-makers. For example, mechanisms of improving representation like deliberative mini-publics that can look very similar to citizen forums in science governance (Farrell, 2014; Setälä, 2017). Deliberative mini-publics decentre the responsibility for policy choices to smaller groups of individuals which aims to bridge institutionalised government structures and publics and thus improve the morality and trust of policy decisions. As with most forms of small group deliberation, mini-publics are criticised for tokenism and lack of impact (Setälä, 2017).

More broadly, decentred governance, as is evident in its name, argues for this decentering of the responsibility for policy choices (Braun & Konninger, 2017; Griggs, Norval, & Wagenaar, 2014). Decentred governance is based on principles of power relations, allowances for pluralistic views in democracy, contestation as central to politics, and shifting decision-making power out of central government (Griggs et al., 2014). Braun and Konninger (2017) call for public engagement to be evaluated through such a lens, specifically that "a decentred analysis of governance approach may be useful for studying participation in science governance in a more complex and holistic perspective without losing sight of power relations, conflict and contestation" (p. 10). Decentred governance allows for a multitude of actors and sites of negotiation of how data is used in government as well as a recognition, similar to feminism, of how power is key to whether publics have influence over decision-making. This aligns somewhat with Galston's (2010) challenge for realism in political theory where "politics is always and everywhere a tension between the drive for and goods of stabilization and consensus, on the one hand, and the drive for and goods of destabilization and conflict, on the other" (p. 396). A view of policy-making as narrative creation and governance as contestation and plurality allows for a critical analysis of how public engagement could be successfully implemented in government as well as a better understanding of the role of data in policy-making.

2.4 Research Methods

This thesis primarily employs qualitative methods following the case-based focus of contextualism and the exploratory nature of the research question. Quantitative methods are used in the final empirical chapter to provide a larger, confirmatory case within the UK. As discussed in the epistemology section Onwuegbuzie and Leech (2005) argue that quantitative and qualitative methods are equally useful for different research problems and are more similar than dissimilar as they fundamentally both 'involve the use of observations to address research questions' (p. 379) as well as processes of verification and triangulation. They argue instead to see research methods as exploratory or confirmatory

techniques, in which both qualitative and quantitative methods can be applied. In relation to the theoretical approach, qualitative research was seen as most appropriate to exploring the research objectives in terms of understanding the power between publics and government as described both in feminist and decentred governance scholarship. The methods are embedded in real contexts and not in lab-based scenarios, emphasising the feminist nature of the work as well as the problem-driven focus of the epistemology. In order to best explore how public engagement is used, how data is used, and how public engagement could be used, it is necessary to work in the contexts in which these problems exist. In other words directly with publics and in local and central governments. Table 2.1 presents a summary of the specific research methodologies employed in each chapter which are further detailed in the chapters themselves.

A range of methods, as presented in Table 2.1, were employed in this thesis for two reasons. First due to the exploratory nature of the research, it was deemed necessary to approach the research objectives from several different lines of enquiry. Each study addresses one of the three objectives and provides insight into one aspect of the research topic. These separate methods are conceptualised as individual lines of discovery that each operate in concert with the others to better understand the overall relationship between publics, government, and data technologies. Rather than testing one single concept, e.g. public engagement operationalisation, this thesis examines several different social phenomena that exist within the overall research question. This reflects the contextualist approach to epistemology. As well, the different methods were chosen in relation to the context of the research question. For example, ethnography was used in Chapter 4 due to the embedded participation of the researcher in helping to design public engagement activities. Similarly, a quantitative approach was taken in Chapter 8 as the research had built sufficient knowledge at that point to hypothesise more fully about the potential relationship of publics, government, and data technologies within the UK. Hence, a mix of exploratory and confirmatory methods were chosen. This reflects the flexible and pragmatic approach to

epistemology. The second reason for the multitude of methods is to ensure that the thesis represents a detailed and varied programme of research training. The PhD process was approached not just as an in-depth research project but also as an opportunity to expand and challenge the skills of the researcher including techniques ranging from highly quantitative like mediation analysis to highly qualitative like ethnography.

Table 2.1: Methodological and analytic techniques of thesis chapters

Chapter	**Method**	**Analytic Technique**	**Exploratory or Confirmatory**	**Quantitative or Qualitative**	**Objective Addressed**
3	Narrative/critical literature review	Thematic Analysis	Exploratory	Qualitative	1
4	Ethnography and document analysis	Thematic Analysis	Exploratory	Qualitative	1
5	Story completion task	Thematic Analysis	Exploratory	Qualitative	2
6	Semi-structured interviews	Thematic Analysis	Exploratory	Qualitative	2
7	Focus groups and workshops	Thematic Analysis	Exploratory	Qualitative	3
8	Online survey	Tests of difference and regression-based mediation analysis	Exploratory and Confirmatory	Quantitative	3

2.4.1 Ethical considerations of research methods

Both formal and personal ethical dimensions were considered in relation to the research methods. In studies involving interviews, focus groups, and ethnography as well as in online surveys, formal ethical concerns like informed

consent were addressed through information sheets and consent forms. As well data protection was addressed through secure data storage and separation of anonymised transcripts or surveys from participant information. From a broader personal perspective of ethical research practice, this thesis also considered the power between participant and researcher as well as the inherent conflict in qualitative research of maintaining a 'natural' research context while still informing participants of their research participation.

For the former, the topic of the research had minimal personal risk related to safety but did have job-related risks for the participants. Data practices can often be deemed unethical, for example incidents of data loss or concern about not seeking retrospective consent in secondary data analysis. Participants were assured that they would be anonymised, including their place of work and any personal details about themselves including gender. At the start of the research there was a concern that if an individual highlighted incidents of unethical data practice, the researcher would be compelled to report this to the appropriate government officer. If such a major concern came up in any of the qualitative studies, it would be essential that this information would be passed on to a government stakeholder in an anonymous form. However, this concern never came to fruition in the research. In regard to power, the researcher was not, as is a traditional concern, often in a privileged position compared to the participants. In fact, in the ethnographic and interview studies the researcher often had minimal influence over when and how often the research was conducted. As well the researcher was often much younger and more female than the interviewees. While this is not particularly problematic, it is of note that this may limit the access the researcher had to the entirety of the public engagement practice context.

For the latter concern of the 'natural' research context, this was deemed an ethical concern as it was challenging in the ethnographic research to consent every single individual during meetings and large public events. It was deemed appropriate to get a global consent agreement with the organisers of the events

rather than specific consent. This allowed for the context to be maintained, i.e. to not interrupt the meetings or engagement events, and for the main research participants to be aware of their participation. The thesis took a conservative approach to documents included in the analysis, personal emails and correspondences that the researcher had access to were not included in the analysis. In sum, all of the studies were approved by the Research Ethics Board in the Department of Psychology at the University of Bath, and a more personal approach to research ethics was also employed by the researcher outside of these administrative ethics procedures.

2.5 Conclusion

In conclusion, this thesis approaches research from a pragmatic contextualist view of the nature of reality and the nature of knowledge. This results in a flexible and mixed methodological and theoretical perspective that focuses on problem-driven research rather than epistemologically pure methods. A range of methods, both quantitative and qualitative, are used to explore the relationship between government, publics, and data technologies including ethnography, semi-structured interviews, mediation analysis and more. This flexible approach is deemed valuable as it allows the researcher to apply the most useful method and theory to each research objective.

LINKING TEXT

In Chapter 3, results from a critical literature review are presented with the aim to examine the potential for public engagement with government data science through a critical analysis of engagement and trust building in other new technologies. This addresses Objective 1, as it seeks to draw on the history of public engagement with other science and technology areas to create a baseline for the subsequent studies in this thesis. It serves as both a review of the practice of public engagement in general as well as an early opportunity to reflect critically on how public engagement is currently being operationalised on the topic of data science. The term data science is used as it is relevant to the readership of the journal this chapter is being published in, however the engagement exercises reviewed are in the context of data technologies in general and not necessarily specific to data science. This paper is both an introduction to the literature used throughout the thesis and an empirical chapter that examines recommendations for the spaces for publics in government data practice. This chapter has been accepted for publication at the journal Government Information Quarterly.

CHAPTER 3

How should we do engagement? A literature review of public engagement for government data science

CHAPTER DECLARATION

This declaration concerns the article entitled:									
How should we do engagement? A literature review of public engagement for government data science									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**		**Accepted**	✓	**Published**	
Publication details (reference)	Rempel, E.S., Barnett, J., and Durrant, H. (2018). How should we do engagement? A literature review of public engagement for government data science. *Government Information Quarterly.* (in press).								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate predominantly executed the formulation, design, work and presentation of the data. Formulation of ideas: ESR designed 85% of the idea including identifying the research question and purpose, while JB and HD contributed the remaining 15% via supervision. Design of methodology: ESR designed the search methodology (85%) in consultation with JB and HD (15%). *"Experimental"* work: ESR undertook 90% of the work including the literature searching, paper identification, and thematic analysis. JB and HD contributed to the remaining 10% by helping with paper identification and refining the thematic analysis. Presentation of data in journal format: ESR wrote up the work for publication including adding reviewer edits (85%) with editing and supervision provided by JB and HD (15%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							**Date**	01/06/2018	

3.1 INTRODUCTION

While data is nothing new, in recent years there has been a revolution in the mechanisms of creating, collecting, processing, connecting, and applying data. In the United Kingdom, the recent passing of the Investigatory Powers Act (Investigatory Powers Act 2016, 2016; MacAskill, 2016) evidences that government, in particular, is attentive to these new possibilities that data affords. Indeed, it has a large stake in the development and use of 'data science'. Traditionally government is positioned as a regulator of new technologies such as genetic modification and nanotechnology (M Kearnes, Macnaghten, & Wilsdon, 2006; Stilgoe, 2007), however it is both the producer and consumer of data science. This metaphorically muddies the waters since data science both informs government policy yet must also be regulated by government. Whilst private sector corporations tend to use data science to expand their business interests (N.A., 2016), governments use data relating to citizen activity to inform decision making across domains ranging from the collection of household waste to tracking terrorist activity (MacAskill, 2016).

A range of recent high profile events such as Care.Data (Carter, Laurie, & Dixon-Woods, 2015) have shown that government data use can be judged as unacceptable, and in the case of Edward Snowden revealing unprecedented invasions of privacy and civil rights (Greenwald, MacAskill, & Poitras, 2013). This also demonstrates how individuals are uniquely tied to data science as a technology due to them often providing the data that government use. Building on this, the argument that this paper advances is that the precarious basis for government self-regulation heightens the importance of including diverse public voices in the realm of data science. Therefore, this paper contributes five propositions developed from a thematic synthesis of public engagement literature for how and why government should include these public voices. The focus of these propositions is on government activities, however there are of course many ways in which publics self-organise and themselves seek conversations on data science with government (Braun & Konninger, 2017; Lezaun & Soneryd, 2007;

Selin et al., 2017). While we do not offer propositions for these publics, we recognise that the high-profile events mentioned above are often driven by publics and subsequently open spaces for these same publics to take part in government activity.

In the light of previous cases of public disaffection around the introduction of new technology, as well as early examples of publicity around data being shared in ways that breach public expectations, it is unsurprising that government has started to initiate public engagement processes around data science. Broadly engagement is defined as the inclusion of publics in some aspect of the development or regulation of policy or technology (Rowe & Frewer, 2005). Rowe and Frewer (2005) categorise public engagement as communication, consultation, or participation. In these three categories the nature of public involvement changes from informing publics about some aspect of a technology to the participation of publics in the deliberation and development of technology. In an alternative classification pertaining more to the motivations of the instigator than to the methods used to engage, Fiorino (1990) and later Stirling (2004) describe engagement as either normative, substantive, or instrumental. Under a normative perspective publics are seen as having a democratic right to engagement in relation to a new technology, for example by living in an area to be affected by its siting. A substantive perspective positions publics as able to make a difference by virtue of their engagement. For example, publics could be involved in developing a new transport system and provide insight about features that would be most valued. Adopting an instrumental perspective would include publics as a means to an end (Barnett, Cooper, & Senior, 2007), for example as a mechanism of minimising controversy or unwanted media attention. While these are useful frameworks, there is no prominent theory of public engagement and thus engagement encapsulates a broad range of public-technology interactions. We define public engagement as a subset of democratic activity that focuses specifically on the inclusion of non-technical publics in the development and governance of new technologies.

We seek to inform our consideration of the potential for public engagement with data science through a critical analysis of engagement and trust building in other new technologies. In the background, we first establish definitional clarity on data science before outlining the interface between data science and the citizenry. We then consider some early examples of public disquiet around government data practices. We then outline the methodology employed in the narrative literature review. In the first half of the discussion, we outline five themes drawn from literature around new technology and public engagement. The second half of our discussion develops these themes into five propositions for data science public engagement in government. We also include examples of how government has begun to engage around data science, as well as how we may use the tenets of decentred governance for theoretical reflection on these propositions.

3.2. BACKGROUND

3.2.1 Definitions of data science

While traditionally defined as the processes of combining and applying data, we use the term data science to represent *both* the substance and application of data. This includes both novel types of data like big data and novel applications of data. Also included within data science is the concept of open data which refers to data sets that are released in the public domain (Hand, 2013). Definitions of big data are more contentious with early classifications relating to particular characteristics, famously first as the three Vs of velocity, volume, and variety (McNeely & Hamh, 2014) but it has also been defined according to its political and social implications by Markus and Topi (2015, p. 3) who call for "a sociotechnical perspective, viewing Big Data as a cluster or assemblage of data-related ideas, resources, and practices." We consider data science to include both the technical practices of data usage and data technology development, but also the ways that data science interacts with, informs and is informed by social and political practice. Data science is a term typically used in the UK, thus the literature we draw on and the public engagement propositions developed are most relevant to the UK context. We include some international literature, mostly from the US, but

do not claim that this study can directly translate to international contexts. It is vital to be clear that our focus is on how publics may engage with the regulation and potential applications of data science and not public engagement using data. We are not focussing on how data technologies can be used as a *means* of engagement, for example government using social media for communication or citizens using social media to organise protests (Lee & Kwak, 2012; Warren, Sulaiman, & Jaafar, 2014), but rather how public engagement about data science and its various contingencies can be done.

3.*2.2 Governments, citizens, and the future of data science*

From Twitter feeds and sentiment analysis, to store card data and marketing, data and data science are claimed to hold vast potential to improve the efficiency of government processes and enhance policy development (Gov.UK Blog, 2015). Complex analytics are being used to predict likely crime scenes in Los Angeles while Facebook 'likes' have been piloted for public health monitoring in Florida (Gittelman et al., 2015; PredPol, 2015). The UK Department for Education recently proposed a register in England which would link diverse data on every school and student (Vale, 2016) while, the Investigatory Powers Act (2016) allows for full-scale web-scraping of all citizens' Internet use (MacAskill, 2016). As Mai (2016, p. 2) describes "what is done in the private sphere and what is done in the public sphere becomes almost impossible to distinguish." There is not only a proliferation of data but of applications through which government can surveil and ultimately regulate the citizenry. And as highlighted through the Snowden case, these applications are not always deemed to be in the public interest.

The routines, interactions, and practices of citizens are inevitably intertwined – albeit often unknowingly - with the production and use of data science. As a citizenry our actions are translated into data in a multitude of ways (Johnson, 2014). For example, by directly filling out administrative forms and clicking links online but also through entering a car park that uses ground sensors. Governance and policy-making for health, employment, banking, transport,

education, justice, and housing, to various degrees, involve the production of data that can in theory contribute to better outcomes for citizens and government (Joseph & Johnson, 2013). Such outcomes might include enhancing how we interact with various organisations which can in turn realise benefits of saving time or money (Hancock, 2016). However, our relationships with data science may be less passive, for example utilising online information about housing to decide where to live through to interactive mapping exercises like Map Kibera that take advantage of local knowledge to overcome shortfalls in government transparency (Donovan, 2012). In this sense, data science, government, and the citizenry exist in a complex data ecosystem with varying, and perhaps decreasing in the case of the citizenry, levels of power and influence. A key message from these tightly intertwined processes is that publics have a fundamental stake in the results and development of data science techniques, whether that is in the improvement of that technology or the regulation of its use.

3.2.3 Early evidence of public disquiet around government use of data

In common with the introduction and development of a range of technologies, public disquiet and disaffection with the use of public data by governments has coalesced around several high-profile incidents in recent years. In late April 2016, New Scientist reported that the Google-owned machine learning firm DeepMind had unprecedented access to all patient records from London hospitals run by the NHS Royal Free Trust (N.A., 2016) resulting in concern with regard to the unconstrained sharing of private medical records. In 2014, following concerns around the sharing of personal medical information, the NHS data strategy Care.Data was cancelled (Boseley, 2016). Publics are often uncomfortable with commercial access to private data and prefer data science that has a clear element of public good (Cameron, Pope, Clemence, & Ipsos MORI Social Research Institute, 2014; Davidson et al., 2013; Ipsos MORI, 2016). This suggests interest from publics in the kinds of data that government collect and how they are ultimately used. Hence, data science has entered in to conversations around public engagement with data science acceptability, privacy, and consent

(Gov.UK Blog, 2015; Sciencewise, 2012, 2014). In light of these early examples of public concern, as well as the precarious basis for government self-regulation of data science, we argue for a more nuanced study of public engagement in data science that moves beyond privacy and consent. We advocate, of course, for transparency on what government does with data but the answer to the relatively simple question of whether individuals should know what is happening to their data is assumed to be yes. Publics should know what happens to their information, particularly when data science based technologies that are developed from that data have the potential to do harm. Data is a sociotechnical object and there are critical questions around the interactions between society, government, and data science that need to be queried and deliberated by publics. Thus, we now move on to discuss how the history of public engagement with new technologies can help to develop models for public engagement with government data science.

3.3 METHODOLOGY

Our aim is to examine the potential for public engagement with data science through a narrative literature review of public engagement with other technologies. A narrative literature review focuses on critically developing new models by synthesising the 'most significant items in the field' (Grant & Booth, 2009, p. 94), and does not formally assess the quality of the literature included. A narrative review was deemed most appropriate due to the broad nature of the research question (e.g. any new technology and public engagement of any form) as well as the desire to include texts that were not empirical in the review. We consider Greenhalgh, Thorne & Malterud's (2018, p. 2) distinction between systematic and narrative reviews as "problems that require data…and those that require clarification and insight" respectively. As our question was exploratory in nature, i.e. developing insight, and we were not interested in testing hypotheses, i.e. statistical analysis, from the included literature we chose a narrative review methodology.

Table 3.1: Search terms and databases

Search terms used	
Public	Public* OR People* OR Societ* OR Communit* OR Populace OR Citizen* OR Person* OR Patient*
Engagement	Engagement OR Consultation OR Dialogue OR Involvement OR Discourse OR Participation OR Communicaton[1]
	Risk* OR Perception* OR Acceptance* OR View* OR Opinion* OR Knowledge OR Attitude* OR Awareness OR Impression* OR Viewpoint*
Trust	Trust* OR Confidence
New Technology	"Emerging *Technolog*" OR "New *Technolog*" OR "Modern *Technolog*" OR "Novel *Technolog*" OR "Developing *Technolog*" OR "Rising *Technolog*"
Databases searched	Scopus, Web of Science, and PsycINFO

1 This word was spelled incorrectly in the original database search.

We extracted relevant literature based on a review of the abstract and title, including both peer-reviewed and grey literature. Search terms used and databases searched are presented in Table 3.1. Relevance was decided based on whether the article focused on any kind of public opinion-seeking or engagement with a new technology. The technology had to involve some kind of science application. We also included literature suggested by the second author based on her expert opinion. Trust was used due to expert advice and its prevalence in early reading around nanotechnology engagement, see Wynne (2006) and Walls et al. (2004). The articles were read for themes common to engagement, we followed Braun and Clarke's (2006) methods of reviewing, coding, categorising, and re-categorising key themes. We extracted common themes from a total of 49 articles. A full list of these 49 articles is available in the section titled Additional Data at the end of the chapter. Articles were included until thematic saturation was reached, i.e. that no new themes were found in new articles. Primary themes were first

developed by the lead author then grouped and revaluated by the team. An initial 12 themes were identified which were then combined and simplified as presented in Table 2. These themes were then used to develop five propositions for public engagement with government data science. These propositions and their connecting theme are presented in Table 3.

We further consider the value of the propositions by then evaluating their relevance to public engagement and governance theory. While Rowe and Frewer (2005) and Fiorino (1990) provide useful frameworks of public engagement, they do not provide a theoretical foundation for thinking about how public engagement, as a concept, fits in government and the governance of technology. In this vein, Braun and Konninger (2017) call for public engagement to be evaluated through a holistic, theoretical lens of decentred governance of science and technology. Engagement is then not about finite, defined public consultations but is underlined by the view of politics "as activities of struggle and conflict concerning the meaning of particular issues that inevitably involve power relations" (pg. 10). They suggest "to take controversy and contestation as a point of departure and study how controversies, publics and issues are brought into being" (pg. 11). Considering the propositions identified in the literature, we evaluate how they can be linked to key tenets and principles of decentred governance. The four tenets we draw on are the recognition of power relations, allowances for pluralism, moving the regulation of technology outside of central government structures, and examining sites of contestation as sites of public engagement (Griggs, Norval, & Wagenaar, 2014).

Table 3.2: Initial and final thematic analysis of public engagement literature results

Initial twelve themes	Five final themes
Stereotypes of the evil market versus the irrational activities in engagement activities.	Conceptions of the public in engagement.
Public understanding leads to success of new technologies.	A knowledge deficit and other early science-citizen interactions.
Transparency will lead to trust.	Trust and trustworthiness in public engagement.
Trust versus robustness as the purpose of engagement.	
Risk is multifaceted.	
Mistrust in technology is actually distrust in government.	
New technologies will have unique public risks, unlike other technologies thus context is important.	
Critiques of the view that people and trust halt technological development.	
Deliberation is conceived as small group discussions.	How aims predict methodologies in public engagement.
Public involvement (participation) will lead to success of new technologies.	
Upstream deliberation is key to public engagement but is often not done in practice.	
Publics and other stakeholders often have unshared socio-technical imaginaries of the future.	Imaginaries of the future in public engagement

3.4 A REVIEW OF PUBLIC ENGAGEMENT WITH NEW TECHNOLOGY

This first half of our discussion introduces the five common themes found in the literature on public engagement with new technology, while in the second half we develop these themes into propositions for government-driven public engagement with data science. Examples of data science engagement and

applications are therefore also presented in Section 5. We include literature around various new technologies that are often applications of scientific principles. In line with Nightingale (2014) in considering technologies, we include both the infrastructure and context around technologies that are in essence part of the technologies themselves. For additional reading on themes in public engagement see Braun and Konninger (2017) and Smallman (2016). First, we discuss conceptions of publics that may be involved in engagement. Second, we outline early initiatives of science-society interactions including the development of the Deficit Model. Next, we discuss lessons for engagement in terms of trust building. Our fourth section describes various methodologies of engagement and how we might move beyond event-based engagement. Finally, we discuss imaginaries of the future and how this impacts the potential for public engagement.

3.4.1 How to define the 'public'

Public engagement, necessarily, involves explicitly or implicitly defining who the public are by deciding who to engage with. Engagement is often about "finding ways of connecting with people who could be mobilised as supporters" (Walker, Cass, Burningham, & Barnett, 2010, p. 942). Various positions are taken relating to whether publics are finite pre-existing groups or constructed and dynamic. Renn (2008) argues for four categories of publics: stakeholders, affected publics, observing publics, and the general public where the general public is the whole or unaffected public. In contrast, Newman (2011) suggests that publics do not pre-exist, rather that public leaders call upon or create publics for a given purpose. Within engagements, the 'real' public is often assumed to be the supportive, silent majority (Burningham, Barnett, & Walker, 2014). In the case of renewable energy technology, Burningham et al. (2014) found publics who oppose new technology were framed as an irrational minority. For example, a common narrative places activists as an unreasonable leftist minority and demonises industry as representing rightist market forces (Barnett, Burningham, Walker, & Cass, 2010; Laurent, 2007; Shelley-Egan & Davies, 2013; Torgersen & Schmidt, 2013). Laurent (2007) similarly found these tendencies to resort to stereotypes were

a common feature of public-industrial interactions around nanotechnology. Government, and other technology stakeholders, often have imaginations of who the public are and what they may feel prior to starting any form of engagement. Furthermore, the mental models that are held of publics and the attributions made about their interests, capabilities, and likely behaviours will determine the nature and extent of the engagement opportunities that are provided (Barnett et al., 2010). Thus, defining 'the public' also defines the spaces for public engagement.

3.4.2 A knowledge deficit and other early science-citizen interactions

The early history of public-technology interactions, later termed public engagement, were framed around the deficit model. This model posits that citizens have a deficit of knowledge around science and technology, and that lack of knowledge drives rejection of technology and scientific ideals. Sturgis and Allum (2004, p. 56) suggest "[implicit] in this programmatic agenda is the claim that 'to know science is to love it'." The deficit model emerged from early science-citizen interactions around technologies such as nuclear power and bioactive substances like thalidomide (Wynne, 2006). In the case of thalidomide, public concerns arose from associations of a causative link between a particular drug and birth defects with more general concerns around science and technology. However, in relation to technologies such as nuclear power, there was an underlying assumption that concern arose from a lack of understanding and knowledge (Grove-White et al., 2004; Wynne, 2006). The "assumption was that no rational and properly informed person could possibly disagree with the desirability of whatever science endorsed – nuclear power, chemical pesticides, chlorofluorocarbons" (Wynne, 2006, p. 215). Following this logic, the method to improve public acceptance (and to ensure the unimpeded progress of such technologies) is to educate and inform the public. This synthesis of communication (Rowe & Frewer, 2005) with instrumental engagement (Fiorino, 1990) was seen in initiatives to inform the general public about technologies to thus ensure their acceptance.

While Sturgis and Allum (2004, p. 55) point out that a "scientifically literate citizenry is also one that can effectively participate in public debates about science", critics of the deficit model take issue with the belief that knowledge about technology guarantees acceptance (M Kearnes et al., 2006; Stilgoe, 2007; Stilgoe, Irwin, & Jones, 2006; Wilsdon, Wynne, & Stilgoe, 2005; Wynne, 2006). As Williams, Macnaghten, Davies, and Curtis (2015, pp. 98-99) argue in their recent critique of deficit-like assumptions in fracking engagement, there "is no guarantee that more information will lead to greater acceptance, or that the availability of facts will lead to a more 'rational' and calculative form of choice-making." In fact, the model reduces publics to mere receivers of information. As Wynne (2006, p. 215) emphatically characterizes, the deficit model is "dogmatically authoritarian and arrogantly self-centred." Knowledge provision, while a component of public engagement, is not sufficient to engage publics in any significant way.

3.4.3 Trust and trustworthiness in public engagement

In the wake of, and as a response to the dominance of the deficit model, engagement exercises were seen as a method of building trust in new technologies (Burningham et al., 2014; Grove-White, Macnaghten, & Wynne, 2000; Groves, 2011; Kearnes & Wynne, 2007; Marris & Rose, 2010; Stilgoe, 2007; The Global Environmental Change Programme, 1999; Torgersen & Schmidt, 2013; Wilsdon et al., 2005). Trust, or a "firm belief in the reliability, truth, or ability of someone or something" (Oxford University Press, 2017), is often cited as key to successful science-citizen interactions. As Warburton (2009, p. 32) reflected after the failure of a nuclear power consultation, an "atmosphere of hostility, caution and anxiety is not conducive to the flexible and creative environment that is ideal for the design and delivery of engagement activities." A lack of trust is problematic. The role of trust is reflected in the genetically modified organism (GMO) and bovine spongiform encephalopathy (BSE) events of the mid-1990s where publics' concern arose around the potentially harmful effects of genetically modifying crops and poor food safety. Trust was diminished following minimal communication and transparency on uncertainties in these technologies, prompting what the UK

House of Lords (2000) deemed a crisis of trust in science. By the mid-2000s, engagement around trust building for things like nanotechnology was commonplace (Groves, 2011; M Kearnes et al., 2006).

Critics of trust building argue it is premised on the same faulty assumption as the deficit model, i.e. the public are a problem (Kearnes & Wynne, 2007; Stilgoe, Lock, & Wilsdon, 2014). As Groves (2011, p. 792) describes, trust building "sees technology and expertise as threatened by dynamics of distrust that disrupt what is imagined as a pre-existing condition of consensual trust in the promises of science". The assumption is that "the prevailing deficit of public confidence…will be improved through the direct consultation and active engagement [of] lay-public concerns" (Kearnes & Wynne, 2007, p. 133). Public involvement then becomes about filling a deficit of trust (Burgess, 2014; Groves, 2011; Involve, 2015; Ipsos MORI, 2006; Kearnes & Wynne, 2007). A focus on trust can, therefore, be seen as merely another instrumental mechanism of preventing public controversy (Krütli, Stauffacher, Flüeler, & Scholz, 2010; Marris & Rose, 2010). As Stilgoe et al. (2006, p. 20) argue, "[the] focus on trust turns the problem into one of communication. And the deliberate attempt to manufacture trust can look deeply untrustworthy." Public engagements that focus on building trust risk doing the opposite. While trust is a component of public and government interactions, it should not be the aim of public engagement. Similar to communication, trust is necessary but not sufficient.

Stilgoe et al. (2006, p. 21) advocate that "[we] must instead focus on what goes into building trustworthiness". This move from trust to trustworthiness was central to shifting public engagement away from the 'public as burden' argument. Instead of the onus being on publics to gain trust, regulators and developers are instead responsible for inspiring trust. There are several key features of trustworthiness, a common argument being that transparency will lead to trust (Flynn, Ricci, & Bellaby, 2012; Kearnes & Wynne, 2007; O'Hara, 2012; Shelley-Egan & Davies, 2013; Stebbing, 2009). Transparency is public openness in the development and regulation of new technologies, but also in the engagement

process itself, e.g. trust that public voices will in fact be heard and not marginalised during public engagements (Warburton, 2009). This openness is often operationalised as releasing data sets, reporting of minutes, and various other practices of being as clear as possible in how something is developed and conclusions about it are drawn. Stebbing (2009, p. 41), in her review of the potential for nanotechnology engagement, argued for "governance that is based on transparency and accountability at the local rather than global level". Thus, the concepts of transparency and accountability are closely linked. The argument is if governance and development are open then government and industry are accountable for good practice in technology development. While transparency and openness in innovation are components of good practice, particularly in government, transparency rests on the assumption that if the processes behind innovation are more widely publicised, developers *might* do the right thing (Kearnes & Wynne, 2007). Transparency is not a replacement for public engagement, it merely facilitates the potential for successful engagement to take place.

A second feature of being trustworthy is understanding that trust is multifaceted. For example, Walls, Pidgeon, Weyman, and Horlick-Jones (2004) suggest that trust in government agencies is rarely simple, it is a critical trust. Publics have neither blind belief nor total scepticism in government agencies. Instead "perceptions and understandings of government agencies and departments are vague and…susceptible to contingencies of events" (Walls et al., 2004, p. 145). For example, Barnett et al. (2007) found that having a stronger belief in public efficacy was associated with having greater trust in genetic science but conversely lower trust in government. Trust becomes attached to different events, histories, and organisations. Walls et al. (2004, p. 135) call for "a situation whereby trust is provisionally conceptualised as multi-faceted, potentially dynamic, and dependent upon a range of contextual variables." Trustworthy practice requires consideration of these different histories and contexts of individuals and publics

who may interact with new technologies, as well as the complexity of public views on government.

3.4.4 Linking the aims and methodologies of public engagement

If trust and communication are only components of engagement, and not the goal, then what is the aim of public engagement? Groves (2011, p. 787) suggests building robustness, i.e. "produc[ing] forms of technology which are explicitly embedded in different ways of domesticating uncertainty". A focus on robustness aims to encourage public participation that emphasises negotiating technological development and governance and creates feedback mechanisms between stakeholders and publics. This shifts engagement away from 'public as problem' and towards 'public as partner' (Stilgoe, 2007). Robustness aligns, to some degree, with Fiorino's (1990) substantive and normative aims of engagement. Normative engagement positions publics as due a role in technological development while substantive engagement posits that publics can improve technology. Hence, publics can provide a unique perspective on the purposes and design of technologies and thus build social and technological robustness.

The aim of engagement is fundamentally linked to how to engage. For communication or trust building purposes, engagement could be typified by providing information. With a normative aim, engagement could include public workshops and early consultation on public opinion and under a substantive viewpoint engagement could involve working groups that ask publics to help develop and design a new technology. Building robustness requires both substantive and normative methodologies of engagement. Bonney, Phillips, Ballard, and Enck (2016) highlight the importance of offering multiple sites for participation, e.g. offering more than one way and one time period for people to engage. They suggest that publics are empowered through involvement in question development. Most authors agree that so-called upstream deliberation is key to public engagement (Burri & Bellucci, 2008; Grove-White et al., 2004; Jones et

al., 2014; M Kearnes et al., 2006; Pidgeon, Harthorn, Bryant, & Rogers-Hayden, 2009; The Global Environmental Change Programme, 1999; Willis & Wilsdon, n.d.; Wilsdon et al., 2005). Burri and Bellucci (2008, p. 387) highlight the "consensus that the public should be involved in deliberative discussions and assessments of emerging technologies at a much earlier stage of technological developments." However, authors caution that engagement should be at the right time rather than simply early. Stilgoe (2007) suggests that publics should be engaged along the research process. Torgersen and Schmidt (2013, p. 52) argue that early engagements result in "participants [that] are difficult to find and to engage, they often discuss half-heartedly in an artificial setting." While early engagement is important to building robustness, it must not overshadow the more important goal of ensuring that publics are engaged meaningfully. In other words, the how of engagement must reflect the why. If individuals are engaged under a normative or substantive aim, then that engagement must have an impact on technological development.

The vast majority of engagement exercises reviewed in developing this paper were time-limited small group discussions, such as workshops or focus groups (Burri & Bellucci, 2008; Flynn et al., 2012; Ipsos MORI, 2006; Jones et al., 2014; M Kearnes et al., 2006; Pidgeon et al., 2009; Stilgoe, 2007; Walker et al., 2010). Pidgeon et al. (2009) and M Kearnes et al. (2006), for example, conducted workshops that included presentation, discussion, and hypothetical deliberation around nanotechnology. Flynn et al. (2012) held Citizens' Panels around hydrogen energy that used presentation, discussion with experts, and again hypothetical deliberation. In this literature, public engagement is often one-off events. That is not to say that all engagement is still event- based, drawing from policy literature, forms of digital engagement are emerging (Gagliardi et al., 2017; Panagiotopoulos, Al-Debei, Fitzgerald, & Elliman, 2012), an interesting example being Lauriault and Mooney's (2014) descriptions of crowdsourcing mapping and other forms of collaborative work. A clear lesson is that these kinds of more involved, participatory engagements are preferable over the older one-off event-style

engagements, and that both kinds of engagement persist in today's government-citizen interactions.

3.4.5 Imaginaries of the future for data science engagement

Engagement generally falls around two main topics: regulation and development. The former being more common than the latter. For example, Flynn et al. (2012) asked participants to debate the potential for a hydrogen economy and drew out themes on hazards, risks, trust, and regulation. Similarly, Burri and Bellucci (2008, p. 388) asked participants to discuss the potential for nanotechnology with an aim at stimulating debate and "to help decision makers in assessing nanotechnologies." While development is a rarer topic, it can be framed by theories of co-design and participatory research (Cornwall & Jewkes, 1995), where in essence publics are engaged through the technology design process itself. Burgess (2014) describes a hybrid where publics were engaged in discussions around genetic technologies and involved in governance along the research process. However, what is missing in these topics is the very first stage of development, i.e. what technologies are developed and why?

Groves (2011), M Kearnes et al. (2006) and Stilgoe et al. (2006) describe these visions of the future as sociotechnical imaginaries, or narratives of how society envisions the future. These imaginaries create the boundaries for what is and is not acceptable in technological development, and are developed through complex socio-technical relationships between industry, academia, media, publics, and government. Groves (2011) argues that the dominant imaginary is an *empty future horizon*. This empty future is one where anything is possible and if anything is possible then there are no limits on development or developer autonomy. However, if there are no limits then what role is there for publics to shape what the future looks like? Some authors argue that ambivalence around new technology often masks a deeply engrained fatalism in public engagement exercises (Grove-White, Macnaghten, Mayer, & Wynne, 1997; Kearnes & Wynne, 2007). Essentially publics do not feel like they can make a difference. Groves (2011,

p. 792) goes as far as to call upon industry and government to conceive of engagement "as a constitutive part of a democratic rewriting of the contract between strategic techno-science and society." The challenge is to allow engagement to be about negotiation and not consultation or communication. Only by first allowing publics to have input on what the future can look like, can more specific engagement, such as regulation or development occur.

3.5 PROPOSITIONS FOR GOVERNMENT-DRIVEN DATA SCIENCE

In the second half of our discussion, we address how each of the themes identified in the literature can be used to build a proposition for government data science public engagement. In doing so we reflect on current and previous government-based data science engagements as well as how the theory of decentred governance can be applied to better understand the theoretical dimensions of the propositions. The connection between the literature review themes and the propositions is presented in Table 3.

Table 3.3: Literature review lessons and propositions for government data science in the UK

Public Engagement Literature Theme Identified		**Proposition for Government Data Science**
Variations in how to define the 'public'	→	Consider the varied and many 'publics' who may be engaged in government data science.
The focus on the knowledge deficit model in early science-citizen interactions	→	Do not assume providing publics with information on data science initiatives will lead to public acceptance.
A shift towards a focus on trust and trustworthiness in public engagement	→	Determine the contingencies of trust for government data science and public engagement through trustworthy practice
How the aim impacts the methodology of public engagement	→	Design public engagements that incorporate robust, critical, and ongoing deliberation of data science
The importance in considering imaginaries of the future for data science engagement	→	Ensure holistic public participation that moves beyond privacy and consent

3.5.1 Proposition One: Consider the varied and many 'publics' who may be engaged in government data science

While identifying the potential publics in any data science project ultimately rests on the specifics of the project itself, it is nevertheless important to consider the potential publics that one may engage with. As with any engagement exercise there are both vast and discrete potential publics, however in practice it is possible to find a subset of the public with whom it is more relevant to engage. Therefore, policy-makers initiating public engagement will benefit from reflecting on how to best characterise and access relevant publics in line with their specific purpose. Data science can also be used to segment populations to identify previously uncategorised groups, for example children in need of educational assistance or individuals most likely to default on a loan (Ginnis et al., 2016; Joseph & Johnson, 2013). And while there are positive and negative aspects to identifying and grouping individuals in this way, it nevertheless requires critical reflection on how to ensure these new publics are both aware of and participants in data science. There are increasingly publics that are already organised and self-identified in relation to data science. Civic hacking groups, in essence public professionals, and associated hackathons offer, in theory, a resource for citizens with non-technical skills to engage with data science. For example, a local civic data group in the UK, Bath: Hacked, recently involved a group of 23 volunteers in a public exercise to create online accessibility maps. Similarly, Lauriault and Mooney (2014) describe a range of group mapping exercises from crowdsourcing to more intensive citizen science. Considering the many and varied publics allows for a range of different opinions to come forward on data science, this is closely linked to the decentred governance tenet of pluralism. Pluralism does not require a consensus from these groups, rather it allows for the reality of the many different ways of looking at an issue. It is also important to note that not all publics are equally powerful in the context of government activities. Non-technical publics and non-governmental publics typically do not have control over the data that is collected from and used to govern them. Government must,

therefore, include a range of publics in data science engagement to fully capture the pluralism of data science governance and challenge problematic power relations between those who collect data and those who are collected upon.

3.5.2 Proposition Two: Do not assume providing publics with information on data science initiatives will lead to public acceptance

Events related to Care.Data illustrate that discussions around government data science were predicated on a deficit model. Care.Data was a data sharing programme that proposed linking general practice records across England. In early 2013, NHS England ran a leaflet information campaign to inform the public of the Care.Data initiative. The NHS assumed that providing knowledge would be sufficient to establish, as Carter et al. (2015) describe, a social license to reuse medical data. However, the poor quality of the information provided, lack of public consultation, and unclear opt-out mechanisms led to public and media concern and many patients requesting to opt out of any sharing of their medical records (Carter et al., 2015; Kirby, 2014). Subsequently, the entire programme was abandoned (Boseley, 2016). Care.Data demonstrates the faulty assumptions around knowledge provision and public acceptance, as well as providing a cautionary tale around government self-regulation. Further to this highly publicised case, a public and professional consultation on data sharing conducted by Cameron et al. (2014, p. 5) suggested that "many participants were sceptical of the value of informing the general public about the ADRN [administrative data linkage] initiative" citing that participants felt publics "would not understand such a complex topic through simple messages, and thus would become worried about data security and privacy when there is not necessarily a reason to be." The notion that even communication is not needed in data initiatives is concerning. We link these deficit-like assumptions to power relations in decentred governance. Assuming that informing publics is enough to gain acceptance, is if nothing else, a reflection on the relative power between government, data scientists, and publics where publics are not provided with the option to critique. No matter how laudable data sharing initiatives may be, government must allow for spaces of

public participation that recognise these power imbalances. This also demonstrates a form of public engagement where publics self-organised around an issue to contest government practice. This is both an example that exposes the imbalance of power between government and publics while also demonstrating publics reasserting their call for power in data science governance (Braun & Konninger, 2017). And while there is recognition in the literature that citizens need decision-making power in data-facilitated engagements like e-participation platforms, (Attard, Orlandi, Scerri, & Auer, 2015; Gagliardi et al., 2017), in practice engagement on data governance is still limited to communication.

3.5.3 Proposition Three: Determine the contingencies of trust for government data science and public engagement through trustworthy practice

Debates around trust, accountability, and transparency have also begun in the realm of data science. One key form of new data, open data, rests on the ideal of transparency (Levy & Johns, 2016; O'Hara, 2012; Schrock, 2016). While open data covers any form of publicly accessible dataset, O'Hara (2012, p. 4) argues that it specifically offers "the possibility of holding government accountable." A key focus in the UK government's recently announced algorithmic decision-making inquiry is "how algorithmic decision-making can be conducted in a 'transparent' or 'accountable' way" (Commons Select Committee, 2017). Transparency in data practices is connected to ideals of accountability while also enabling people to engage with data science (Attard et al., 2015). However, this assumes publics have the resources, skills, and finances to use and interpret the data provided. Arguably, with highly specialist technologies like data science, this is often not the case (Levy & Johns, 2016). In practice, as Attard et al. (2015, p. 414) reflects, while "the benefits of open data outweigh the efforts required, it appears there is a lack of public participation in open government initiatives." The open data to accountability perspective has largely been critiqued as being naïve, in fact Johnson (2014) argues that 'opening up data can function as a tool of disciplinary power' (pg. 270) and further suggests open data theorists move to a perspective of *information justice*, which 'can be especially useful in overcoming the capabilities

gaps between enterprises and individuals…and make information pluralism a reality (p. 272). While not an argument for public engagement per se, it is a perspective that allows for the socio- to reintegrate to technical systems.

In terms of considering the various contingencies and contexts of trust in data science, it is challenging to divorce views of government from views of government data practice. In a public consultation on data linking, Cameron et al. (2014, p. 14) found that "participants trusted government's intentions more than commercial companies" and yet "were also worried about personal data being leaked, lost, shared or sold by government departments to third parties." They concluded that "[l]ow trust in government more generally seemed to be driving these views" (Cameron et al., 2014, p. 22). It is challenging to distinguish what publics may use as reference points to evaluate data risks. Cameron et al. (2014) suggest that media events, like the revelations of Edward Snowden and Julian Assange, drive the belief that data can never be truly secure. In a widely publicised case in the UK, Google subsidiary DeepMind developed a collaboration agreement with the NHS Royal Free Trust gaining access to millions of identifiable patient records. While they claimed to only be accessing these records to provide an app that identifies acute kidney (AKI) patients, as Powles and Hodson (2017) discuss there were in fact no real limitations on what they could do with the data, nor any transparency in what they were actually doing. In addition, patients were enrolled based on a principle of implied consent due to the app being involved in direct care, beyond being a shaky assumption on any grounds, implied consent would only be in effect for those patients with AKI (Powles & Hodson, 2017). After the New Scientist revealed the problematic latitude of this agreement (N.A., 2016) Google and DeepMind published press releases highlighting the small scope of the project in order to reassure publics. Despite, or perhaps due to, these minimal *post hoc* attempts at public participation this data science collaboration remains under scrutiny.

It is events like these, where untrustworthy practice comes to light, that are likely to drive public perceptions of data science. UK governments have made

some efforts to define how publics feel about data. Examples include the Cabinet Office's Government Digital Service consultation on what the 'red lines' are in ethical data science (Gov.UK Blog, 2015; Sciencewise, 2012, 2014). In 2015, the Office for National Statistics and the Wellcome Trust held a series of deliberative workshops with general publics and specialist groups to identify how people felt about commercial access to health data (Ipsos MORI, 2016). Government-commissioned reports on public views of data have been implemented or funded by the Economic and Social Research Council , Ipsos MORI , the Scottish Government, Sciencewise, the Cabinet Office and more (Cameron et al., 2014; Davidson et al., 2013; Gov.UK Blog, 2015; Ipsos MORI, 2006, 2016; Sciencewise, 2012, 2014). While these activities represent a step toward establishing publics' views on various aspects of data science, they do not equate nor reflect public influence. Despite how they may be designed around finding the ethical lines of data science, these are not sufficient for public engagement nor do they overcome untrustworthy practice in other areas. In fact these consultations evidence a lack of any kind of decentred governance. Particularly in the Google DeepMind case, they evidence attempts to 'get around' public engagement rather than critical inward reflections on trustworthy practices that could enable public engagement. Thus the main tenet of decentred governance that regulation of data science would occur beyond central government can only be built upon trustworthy practice.

3.5.4 Proposition Four: Design public engagements that incorporate robust, critical, and ongoing deliberation of data science

As data is already intertwined with publics, creating more robust models of public ownership of personal data could move governments beyond simply conducting one-off events. The ongoing embedded nature of public data practices calls for designing innovative engagement exercises that reflect how data science is increasingly a part of day-to-day life. Opportunities for involving publics in data collection, use, and governance abound. Examples include citizen science initiatives that crowd-source data processing (Bonney et al., 2016; Lauriault & Mooney, 2014) and the aforementioned hackathons that leverage the knowledge of

public professionals in typically government-led events to drive data innovation (Sousa, 2013). Historical exercises in the US like citizen engagement through commenting and rule-making could, of course, extend to data policies and practices in the UK. These efforts allow for open public comments on policy drafts, see Johns and Saltane (2016). However, there is the potential for more inclusive forms of engagement with data than these limited one-off events. Innovative forms of engagement using social media are already in practice (Lee & Kwak, 2012), however it is important to note that these are often engagements using data technologies and not about data science itself.

There is also an important lesson in thinking critically about the aim of these data science engagements. Substantive forms of engagement should have the possibility to make a difference and to enable publics to have impact on the processes of data science. As Malik (2013, p. 6) describes, "[the] first step of the journey toward Big Data governance involves stakeholder engagement". As government is facing increasingly complex challenges in how to regulate their own use of data science technologies, innovative and multiple venues for public participation can help government address broader questions around ethical and beneficial data science. The technical knowledge required for conducting data science is obviously also a limitation, and thus focussing on ethical questions would allow broader engagement. This will require new and innovative forms of public engagement that allow for sites of contestation and pluralism, this 'means that participation in science governance takes place in a multiplicity of sites' (Braun & Konninger, 2017, p. 10). And while it is impossible to predict where these sites of contestation will develop, public engagement that is robust, critical, and ongoing allows public engagement to be less about shutting down public opposition and more about opening up debate (Stirling, 2004).

3.5.5 Proposition Five: Ensure holistic public participation that moves beyond privacy and consent

What is the future imaginary of data? Thus far the topics of data science engagement have focussed on privacy and consent (Joseph & Johnson, 2013; Schintler & Kulkarni, 2014; Stough & McBride, 2014). For example, Cameron et al.'s (2014) Dialogue on Data sought "to explore public understanding and views of administrative data and data linking". Limiting public engagement to discussions around privacy and consent sets a tight boundary for public influence and can be seen as a way of shutting down potential areas for conversation and contestation. We suggest a future horizons perspective where there is an opportunity for government to involve publics in creating an imaginary of the data future. More than upstream deliberation or building robustness, government can think about the way publics can be involved in the ideology that drives the use of data science in policy, and the kinds of data that are created. This can be achieved through more inclusive and early engagement. As Kennedy, Poell, and van Dijck (2015, p. 6) argue, '[to] participate in datafied social, political, cultural and civic life, ordinary people need to understand what happens to their data, the consequences of data analysis, and the ways in which data-driven operations affect us all.' We would go further to add that 'ordinary people' need to be understood as the key stakeholder in a datafied world. Publics need to be engaged in how they imagine data, what sorts of information they see as useful, how they think data could be used, and most importantly in how they wish the future to look. Only through this kind of decentralized governance with data science can the issue of government self-regulation be addressed.

3.6 CONCLUSION

Data science public engagement is a burgeoning field of practice within government and academia. However, there is much to learn from public engagement in other fields that can be applied to this practice. We conclude by reflecting on the strengths, limitations and academic contributions of this paper as well as the overall conclusions.

3.6.1 Strengths, limitations, and academic contribution

This is the first article to offer propositions for government data science public engagement practice that are rooted in concrete and empirical lessons from a narrative review of the public engagement literature. This is the main academic contribution of this paper. In addition, we have reflected on past data science engagement exercises as well as possible future directions. We suggest both academic scholars and government officials could build on these lessons to develop effective public engagement activities. This article focuses on introducing the reader to the field of public engagement and its application to data science. While we did not intend to conduct a systematic review of the literature, lessons found here could be strengthened and enhanced by further reviews of public engagement in the future. A more critical and theory-driven approach could usefully inform more in-depth examination of particular domains of data or data science practice.

3.6.2 Final thoughts

The increasing use of data science in policymaking is creating new spaces for public engagement. These new opportunities can create confusion on how and where to effectively engage publics in the development and regulation of data science. We synthesised five themes from 49 articles that focussed on public engagement with new technology. These themes were then used to develop five novel propositions for public engagement with government data science. This includes considering the varied and many 'publics' who may be engaged in government data science, not assuming that providing publics with information on data science initiatives will lead to public acceptance, determining the contingencies of trust for government data science and public engagement through trustworthy practice, and designing public engagements that incorporate robust, critical, and ongoing deliberation of data science. Our final proposition is to ensure holistic public participation that moves beyond privacy and consent. This highlights the importance of recognising that publics have an interest in *how* and *why* government uses data science. In particular they are due a role in deciding what government *should* use data science for. Government has a unique

opportunity to allow publics in decision-making spaces around how data is created, collected, and utilised for the good of society. Data is a public matter. It is the next steps that government take that will decide whether publics are adversaries or partners in this data future.

REFERENCES

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*(4), 399-418. doi:10.1016/j.giq.2015.07.006

Barnett, J., Burningham, K., Walker, G., & Cass, N. (2010). Imagined publics and engagement around renewable energy technologies in the UK. *Public Understanding of Science, 21*(1), 36-50. doi:10.1177/0963662510365663

Barnett, J., Cooper, H., & Senior, V. (2007). Belief in public efficacy, trust, and attitudes toward modern genetic science. *Risk Analysis, 27*(4), 921-933. doi:10.1111/j.1539-6924.2007.00932.x

Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Underst Sci, 25*(1), 2-16. doi:10.1177/0963662515607406

Boseley, S. (2016). NHS to scrap single database of patients' medical details. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2016/jul/06/nhs-to-scrap-single-database-of-patients-medical-details

Braun, K., & Konninger, S. (2017). From experiments to ecosystems? Reviewing public participation, scientific governance and the systemic turn. *Public Underst Sci*, 963662517717375. doi:10.1177/0963662517717375

Burgess, M. M. (2014). From 'trust us' to participatory governance: Deliberative publics and science policy. *Public Underst Sci, 23*(1), 48-52. doi:10.1177/0963662512472160

Burningham, K., Barnett, J., & Walker, G. (2014). An Array of Deficits: Unpacking NIMBY Discourses in Wind Energy Developers' Conceptualizations of Their Local Opponents. *Society & Natural Resources, 28*(3), 246-260. doi:10.1080/08941920.2014.933923

Burri, R. V., & Bellucci, S. (2008). Public perception of nanotechnology. *Journal of Nanoparticle Research, 10*(3), 387-391. doi:10.1007/s11051-007-9286-7

Cameron, D., Pope, S., Clemence, M., & Ipsos MORI Social Research Institute. (2014). *Dialogue on Data - Exploring the public's views on using linked administrative data for research purposes*. Retrieved from London, UK: https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics, 41*(5), 404-409. doi:10.1136/medethics-2014-102374

Commons Select Committee. (2017). Algorithms in decision-making inquiry launched. Retrieved from https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2015/algorithms-in-decision-making-inquiry-launch-16-17/

Cornwall, A., & Jewkes, R. (1995). What is participatory research? *Social Science and Medicine, 41*(12), 1667-1676.

Davidson, S., McLean, C., Treanor, S., Ipsos MORI, Aitken, M., Cunningham-Burley, S., . . . University of Edinburgh. (2013). *Public acceptability of data sharing between the public, private and third sectors for research purposes*. Retrieved from

Donovan, K. (2012). Seeing like a slum: Towards open, deliberative development. *Georgetown Journal of International Affairs, 13*(1), 97-104.

Fiorino, D. J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. *Science Technology Human Values, 15*(2), 226-243.

Flynn, R., Ricci, M., & Bellaby, P. (2012). Ambiguity, complexity and uncertainty surrounding the hazards of hydrogen and public views of emergent risks. *Journal of Risk Research, 15*(4), 373-387. doi:10.1080/13669877.2011.634517

Gagliardi, D., Schina, L., Sarcinella, M. L., Mangialardi, G., Niglia, F., & Corallo, A. (2017). Information and communication technologies and public participation: interactive maps and value added for citizens. *Government Information Quarterly, 34*(1), 153-166. doi:10.1016/j.giq.2016.09.002

Ginnis, S., Evans, H., Boal, N., Davies, E., Palmqvist Aslaksen, A., Ipsos MORI Social Research Institute, . . . Sciencewise. (2016). *Public dialogue on the ethics of data science in government*. Retrieved from London, UK: https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf

Gittelman, S., Lange, V., Gotway Crawford, C. A., Okoro, C. A., Lieb, E., Dhingra, S. S., & Trimarchi, E. (2015). A new source of data for public health surveillance: Facebook likes. *Journal of Medical Internet Research, 17*(4), e98. doi:10.2196/jmir.3970

Gov.UK Blog. (2015). Data Science ethics. Retrieved from https://data.blog.gov.uk/2015/12/08/data-science-ethics/

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J, 26*(2), 91-108. doi:10.1111/j.1471-1842.2009.00848.x

Greenhalgh, T., Thorne, S., & Malterud, K. (2018). Time to challenge the spurious hierarchy of systematic over narrative reviews? *European Journal of Clinical Investigation, epub: 2018;e12931, 1-6.*

Greenwald, G., MacAskill, E., & Poitras, K. (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelation. *The Guardian*.

Griggs, S., Norval, A. J., & Wagenaar, H. (2014). Introduction : Democracy, conflict and participation in decentred governance. In S. Griggs, A. J. Norval, & H. Wagenaar (Eds.), *Practices of Freedom: Decentred Governance, Conflict and Democratic Participation*. New York, NY: Cambridge University Press.

Grove-White, R., Kearnes, M., Miller, P., Macnaghten, P., Wilsdon, J., & Wynne, B. (2004). *Bio-to-Nano? Learning the lessons, interrogating the comparison (amended version)*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.658.5110&rep=rep1&type=pdf

Grove-White, R., Macnaghten, P., Mayer, S., & Wynne, B. (1997). *Uncertain world: Genetically modified organisms, food and public attitudes in Britain*. Retrieved from Lancaster, England: Centre for the Study of Environmental Change.

Grove-White, R., Macnaghten, P., & Wynne, B. (2000). *Wising up: The public and new technologies*. Retrieved from Lancaster University: Centre for the Study of Environmental Change.

Groves, C. (2011). Public engagement and nanotechnology in the UK: restoring trust or building robustness? *Science and Public Policy, 38*(10), 783-793. doi:10.3152/030234211x13070021633440

Hancock, M. (2016). *Data science ethical framework launch: Matt Hancock speech*. London: Cabinet Office Retrieved from https://www.gov.uk/government/speeches/data-science-ethical-framework-launch-matt-hancock-speech.

Hand, D. J. (2013). *Data, not dogma: Big data, open data, and the opportunities ahead.* Paper presented at the Advances in Intelligent Data Analysis XII, London UK.

Investigatory Powers Act 2016. (2016). *Investigatory Powers Act 2016*. Retrieved from United Kingdom:

Involve. (2015). *Written evidence submitted by Involve (BIG0010)*. London, UK.

Ipsos MORI. (2006). *The use of personal health information in medical research general public consultation: Final report*. Retrieved from London, UK: https://mrc.ukri.org/documents/pdf/the-use-of-personal-health-information-in-medical-research-june-2007/

Ipsos MORI. (2016). *The one-way mirror: Public attitudes to commercial access to health data*. Retrieved from London, UK: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Johns, M., & Saltane, V. (2016). *Citizen engagement in rulemaking: Evidence on regulatory practice in 185 countries*. Retrieved from https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Johnson, J. A. (2014). From open data to information justice. *Ethics and Information Technology, 16*(4), 263-274. doi:10.1007/s10676-014-9351-8

Jones, A. R., Anderson, A. A., Yeo, S. K., Greenberg, A. E., Brossard, D., & Moore, J. W. (2014). Using a deliberative exercise to foster public engagement in nanotechnology. *Journal of Chemical Education, 91*(2), 179-187. doi:10.1021/ed400517q

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional, 15*(6), 43-48.

Kearnes, M., Macnaghten, P., & Wilsdon, J. (2006). *Governing at the nanoscale: People, policies and emerging technologies*. Retrieved from London:

Kearnes, M., & Wynne, B. (2007). On nanotechnology and ambivalence: The politics of enthusiasm. *NanoEthics, 1*(2), 131-142. doi:10.1007/s11569-007-0014-7

Kennedy, H., Poell, T., & van Dijck, J. (2015). Data and agency. *Big Data & Society, 2*(2). doi:10.1177/2053951715621569

Kirby, T. (2014). Controversy surrounds England's new NHS database. *The Lancet, 383*(9918), 681. doi:10.1016/s0140-6736(14)60230-0

Krütli, P., Stauffacher, M., Flüeler, T., & Scholz, R. W. (2010). Functional-dynamic public participation in technological decision-making: site selection processes of nuclear waste repositories. *Journal of Risk Research, 13*(7), 861-875. doi:10.1080/13669871003703252

Laurent, B. (2007). Diverging convergences - Competing meanings of nanotechnology and converging technologies in a local context. *Innovation-the European Journal of Social Science Research, 20*(4), 343-357. doi:10.1080/13511610701760804

Lauriault, T. P., & Mooney, P. (2014). *Crowdsourcing: A geographic approach to public engagement*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2518233

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public engagement. *Government Information Quarterly, 29*(4), 492-503. doi:10.1016/j.giq.2012.06.001

Levy, K. E. C., & Johns, D. M. (2016). When open data is a Trojan Horse: The weaponization of transparency in science and governance. *Big Data & Society*, 1-6. doi:10.1177/2053951715621568

Lezaun, J., & Soneryd, L. (2007). Consulting citizens: technologies of elicitation and the mobility of publics. *Public Understanding of Science, 16*(3), 279-297. doi:10.1177/0963662507079371

Lupton, D. (2015). The thirteen Ps of big data. Retrieved from https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/

MacAskill, E. (2016). 'Extreme surveillance' becomes UK law with barely a whimper. *The Guardian*. Retrieved from https://www.theguardian.com/world/2016/nov/19/extreme-surveillance-becomes-uk-law-with-barely-a-whimper

Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *The Information Society, 32*(3), 192-199. doi:10.1080/01972243.2016.1153010

Malik, P. (2013). Governing big data: Principles and practices. *IBM Journal of Research and Development, 57*(3/4), 1-13.

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business: Report on a research agenda setting workshop*. Retrieved from Washington, D.C.:

Marris, C., & Rose, N. (2010). Open engagement: Exploring public participation in the biosciences. *PLoS Biol, 8*(11), e1000551. doi:10.1371/journal.pbio.1000551

McNeely, C. L., & Hamh, J. (2014). The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research, 31*(4), 304-310.

N.A. (2016). Revealed: Google AI has access to huge haul of NHS patient data. *New Scientist*. Retrieved from

https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/#.VyOORsEjCiU.twitter

Newman, J. (2011). Public leadership as public-making. *Public Money & Management, 31*(5), 315-322. doi:10.1080/09540962.2011.598336

Nightingale, P. (2014). *What is technology? Six definitions and two pathologies*. Retrieved from Brighton:

O'Hara, K. (2012). *Transparency, open data and trust in government: Shaping the infosphere*. Paper presented at the 4th Annual ACM Web Science Conference.

Oxford University Press. (Ed.) (2017) Oxford English Dictionary. Oxford: Oxford University Press,.

Panagiotopoulos, P., Al-Debei, M. M., Fitzgerald, G., & Elliman, T. (2012). A business model perspective for ICTs in public engagement. *Government Information Quarterly, 29*(2), 192-202. doi:10.1016/j.giq.2011.09.011

Pidgeon, N., Harthorn, B. H., Bryant, K., & Rogers-Hayden, T. (2009). Deliberating the risks of nanotechnologies for energy and health applications in the United States and United Kingdom. *Nature Nanotechnology, 4*(2), 95-98. doi:10.1038/nnano.2008.362

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*. doi:10.1007/s12553-017-0179-1

PredPol. (2015). Predictive Policing: The predictive policing company.

Renn, O. (2008). Risk Participation. In O. Renn (Ed.), *Risk Governance: Coping with Uncertainty in a Complex World*. London: Earthscan.

Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. *Science, Technology & Human Values, 30*(2), 251-290. doi:10.1177/0162243904271724

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research, 31*(4), 343-348.

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society, 18*(4), 581-599.

Sciencewise. (2012). Public dialogue on data openness, data re-use and data management: Sciencewise co-funded project 2012. Retrieved from http://www.sciencewise-erc.org.uk/cms/public-dialogue-on-data-openness-data-re-use-and-data-management/

Sciencewise. (2014). Big Data: Public views on the collection, sharing and use of personal data by government and companies.

Selin, C., Rawlings, K. C., de Ridder-Vignone, K., Sadowski, J., Altamirano Allende, C., Gano, G., . . . Guston, D. H. (2017). Experiments in engagement: Designing public engagement with science and technology for capacity building. *Public Underst Sci, 26*(6), 634-649. doi:10.1177/0963662515620970

Shelley-Egan, C., & Davies, S. R. (2013). Nano-industry operationalizations of "responsibility": Charting diversity in the enactment of responsibility. *Review of Policy Research, 30*(5), 588-604. doi:10.1111/ropr.12032

Smallman, M. (2016). Public understanding of science in turbulent times III: Deficit to dialogue, champions to critics. *Public Understanding of Science, 25*(2), 186-197.

Sousa, S. (2013). *Lessons learnt from a public–private big data hackathon*. Retrieved from London, UK: https://fvstatic.s3.amazonaws.com/1425646063_0599941001425646063.pdf

Stebbing, M. (2009). Avoiding the trust deficit: Public engagement, values, the precautionary principle and the future of nanotechnology. *Journal of Bioethical Inquiry, 6*(1), 37-48. doi:10.1007/s11673-009-9142-9

Stilgoe, J. (2007). *Nanodialogues: Experiments in public engagement with science*. Retrieved from London, UK: https://www.demos.co.uk/files/Nanodialogues%20-%20%20web.pdf

Stilgoe, J., Irwin, A., & Jones, K. (2006). *The received wisdom: Opening up expert advice*. Retrieved from London, UK: https://www.demos.co.uk/files/receivedwisdom.pdf

Stilgoe, J., Lock, S. J., & Wilsdon, J. (2014). Why should we promote public engagement with science? *Public Understanding of Science, 23*(1), 4-15.

Stirling, A. (2004). Opening up or closing down? Analysis, participation and power in the social appraisal of technology. In M. Leach, I. Scoones, & B. Wynne (Eds.), *Science and citizens: Globalization and the challenge of engagement* (pp. 218-231). London: Zed Books.

Stough, R., & McBride, D. (2014). Big data and U.S. public policy. *Review of Policy Research, 31*(4), 339-342.

Sturgis, P., & Allum, N. (2004). Science in Society: Re-Evaluating the Deficit Model of Public Attitudes. *Public Understanding of Science, 13*(1), 55-74. doi:10.1177/0963662504042690

The Global Environmental Change Programme. (1999). *The politics of GM food*. Retrieved from Swindon:

Torgersen, H., & Schmidt, M. (2013). Frames and comparators: How might a debate on synthetic biology evolve? *Futures, 48*, 44-54. doi:10.1016/j.futures.2013.02.002

UK House of Lords. (2000). *Select committee on science and technology: Science and society*. London, UK.

Uprichard, E. (2013). Focus: Big data, little questions? *Discover Society*.

Vale, R. (2016). The register of schools in England goes into alpha. Retrieved from https://data.blog.gov.uk/2016/11/29/the-register-of-schools-in-england-goes-into-alpha/

Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and socio technical change: Imagined subjectivities of 'the public' and their implications. *Environment and Planning A, 42*(4), 931-947.

Walls, J., Pidgeon, N., Weyman, A., & Horlick-Jones, T. (2004). Critical trust: understanding lay perceptions of health and safety risk regulation. *Health, Risk & Society, 6*(2), 133-150. doi:10.1080/1369857042000219788

Warburton, D. (2009). *Evaluation of BERR's engagement of the public and other interested parties in the future of civil nuclear power in the UK: Final Report*.

Retrieved from London, UK: http://www.sharedpractice.org.uk/Downloads/Nuclear_report.pdf

Warren, A. M., Sulaiman, A., & Jaafar, N. I. (2014). Social media effects on fostering online civic engagement and building citizen trust and trust in institutions. *Government Information Quarterly, 31*(2), 291-301. doi:10.1016/j.giq.2013.11.007

Williams, L., Macnaghten, P., Davies, R., & Curtis, S. (2015). Framing 'fracking': Exploring public perceptions of hydraulic fracturing in the United Kingdom. *Public Underst Sci*. doi:10.1177/0963662515595159

Willis, R., & Wilsdon, J. (n.d.). *From bio to nano and beyond: A progressive agenda for technology, risk and the environment*. Retrieved from London, UK: http://www.green-alliance.org.uk/page_234.php

Wilsdon, J., Wynne, B., & Stilgoe, J. (2005). *The public value of science: Or how to ensure that science really matters*. Retrieved from London, UK.

Wynne, B. (2006). Public engagement as a means of restoring public trust in science--hitting the notes, but missing the music? *Community Genet, 9*(3), 211-220. doi:10.1159/000092659

ADDITIONAL DATA

Appendix 1: References used in narrative literature review thematic analysis

Reference Number	Reference
1	*Abbot, C. (2012). Bridging the gap - Non-state actors and the challenges of regulating new technology. *Journal of Law and Society, 39*(3), 329-358. doi:10.1111/j.1467-6478.2012.00588.x
2	*Ackland, R., Gibson, R., Lusoli, W., & Ward, S. (2010). Engaging With the Public? Assessing the Online Presence and Communication Practices of the Nanotechnology Industry. *Social Science Computer Review, 28*(4), 443-465. doi:10.1177/0894439310362735
3	*Asveld, L., Ganzevles, J., & Osseweijer, P. (2015). Trustworthiness and Responsible Research and Innovation: The Case of the Bio-Economy. *Journal of Agricultural and Environmental Ethics, 28*(3), 571-588. doi:10.1007/s10806-015-9542-2
4	*Barling, D., De Vriend, H., Cornelese, J. A., Ekstrand, B., Hecker, E. F. F., Howlett, J., . . . Top, R. (1999). The social aspects of food biotechnology: A European view. *Environmental Toxicology and Pharmacology, 7*(2), 85-93. doi:10.1016/S1382-6689(99)00009-5
5	Barnett, J., Burningham, K., Walker, G., & Cass, N. (2010). Imagined publics and engagement around renewable energy technologies in the UK. *Public Understanding of Science, 21*(1), 36-50. doi:10.1177/0963662510365663
6	Barnett, J., Cooper, H., & Senior, V. (2007). Belief in public efficacy, trust, and attitudes toward modern genetic science. *Risk Analysis, 27*(4), 921-933. doi:10.1111/j.1539-6924.2007.00932.x
7	*Bebbington, A., & Williams, M. (2008). Water and mining conflicts in Peru. *Mountain Research and Development, 28*(3-4), 190-195. doi:10.1659/mrd.1039
8	Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Underst Sci, 25*(1), 2-16. doi:10.1177/0963662515607406
9	*Boyd, A. D. (2015). Connections between community and emerging technology: Support for enhanced oil recovery in the Weyburn, Saskatchewan area. *International Journal of Greenhouse Gas Control, 32*, 81-89. doi:10.1016/j.ijggc.2014.11.005
10	Burgess, M. M. (2014). From 'trust us' to participatory governance: Deliberative publics and science policy. *Public Underst Sci, 23*(1), 48-52. doi:10.1177/0963662512472160
11	Burningham, K., Barnett, J., & Walker, G. (2014). An Array of Deficits: Unpacking NIMBY Discourses in Wind Energy Developers' Conceptualizations of Their Local Opponents. *Society & Natural Resources, 28*(3), 246-260. doi:10.1080/08941920.2014.933923
12	Burri, R. V., & Bellucci, S. (2008). Public perception of nanotechnology. *Journal of Nanoparticle Research, 10*(3), 387-391. doi:10.1007/s11051-007-9286-7
13	*Cegarra-Navarro, J.-G., Pachón, J. R. C., & Cegarra, J. L. M. (2012). E-government and citizen's engagement with local affairs through e-websites: The case of Spanish municipalities. *International Journal of Information Management, 32*(5), 469-478. doi:10.1016/j.ijinfomgt.2012.02.008
14	*Dankel, D. J., Roland, K. L., Fisher, M., Brenneman, K., Delgado, A., Santander, J., . . . Curtiss, R. (2014). Making Common Sense of Vaccines: An Example of Discussing the Recombinant Attenuated Salmonella Vaccine with the Public. *NanoEthics, 8*(2), 179-185. doi:10.1007/s11569-014-0198-6
15	Fiorino, D. J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. *Science Technology Human Values, 15*(2), 226-243.
16	Flynn, R., Ricci, M., & Bellaby, P. (2012). Ambiguity, complexity and uncertainty surrounding the hazards of hydrogen and public views of emergent risks. *Journal of Risk Research, 15*(4), 373-387. doi:10.1080/13669877.2011.634517
17	Grove-White, R., Kearnes, M., Miller, P., Macnaghten, P., Wilsdon, J., & Wynne, B. (2004). *Bio-to-Nano? Learning the lessons, interrogating the comparison (amended version)*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.658.5110&rep=rep1&type=pdf
18	Grove-White, R., Macnaghten, P., Mayer, S., & Wynne, B. (1997). *Uncertain world: Genetically modified organisms, food and public attitudes in Britain*. Retrieved from Lancaster, England: Centre for the Study of Environmental Change.

19	Grove-White, R., Macnaghten, P., & Wynne, B. (2000). *Wising up: The public and new technologies*. Retrieved from Lancaster University: Centre for the Study of Environmental Change.
20	Groves, C. (2011). Public engagement and nanotechnology in the UK: restoring trust or building robustness? *Science and Public Policy, 38*(10), 783-793. doi:10.3152/030234211x13070021633440
21	Jones, A. R., Anderson, A. A., Yeo, S. K., Greenberg, A. E., Brossard, D., & Moore, J. W. (2014). Using a deliberative exercise to foster public engagement in nanotechnology. *Journal of Chemical Education, 91*(2), 179-187. doi:10.1021/ed400517q
22	Kearnes, M., Macnaghten, P., & Wilsdon, J. (2006). *Governing at the nanoscale: People, policies and emerging technologies*. Retrieved from London:
23	Kearnes, M., & Wynne, B. (2007). On nanotechnology and ambivalence: The politics of enthusiasm. *NanoEthics, 1*(2), 131-142. doi:10.1007/s11569-007-0014-7
24	Krütli, P., Stauffacher, M., Flüeler, T., & Scholz, R. W. (2010). Functional-dynamic public participation in technological decision-making: site selection processes of nuclear waste repositories. Journal of Risk Research, 13(7), 861-875. doi:10.1080/13669871003703252
25	*Kuzdas, C., & Wiek, A. (2014). Governance scenarios for addressing water conflicts and climate change impacts. *Environmental Science & Policy, 42*, 181-196. doi:10.1016/j.envsci.2014.06.007
26	Laurent, B. (2007). Diverging convergences - Competing meanings of nanotechnology and converging technologies in a local context. *Innovation-the European Journal of Social Science Research, 20*(4), 343-357. doi:10.1080/13511610701760804
27	Marris, C., & Rose, N. (2010). Open engagement: Exploring public participation in the biosciences. *PLoS Biol, 8*(11), e1000551. doi:10.1371/journal.pbio.1000551
28	*Ngar-yin Mah, D., & Hills, P. (2014). Participatory governance for energy policy-making: A case study of the UK nuclear consultation in 2007. *Energy Policy, 74*, 340-351. doi:10.1016/j.enpol.2014.08.002
29	*Nowotny, H. (2014). Engaging with the political imaginaries of science: Near misses and future targets. *Public Underst Sci, 23*(1), 16-20. doi:10.1177/0963662513476220
30	Pidgeon, N., Harthorn, B. H., Bryant, K., & Rogers-Hayden, T. (2009). Deliberating the risks of nanotechnologies for energy and health applications in the United States and United Kingdom. *Nature Nanotechnology, 4*(2), 95-98. doi:10.1038/nnano.2008.362
31	*Pinnock, H., Slack, R., Pagliari, C., Price, D., & Sheikh, A. (2007). Understanding the potential role of mobile phone-based monitoring on asthma self-management: qualitative study. *Clinical and Experimental Allergy, 37*(5), 794-802. doi:10.1111/j.1365-2222.2007.02708.x
32	*Quested, G., & Sousa, S. (2013). *Big data: A lot to explore but little to share? Ways towards a regulatory regime for big data.* Paper presented at the Big Data Week 2013, London.
33	*Reinsberger, K., Brudermann, T., Hatzl, S., Fleiß, E., & Posch, A. (2015). Photovoltaic diffusion from the bottom-up: Analytical investigation of critical factors. *Applied Energy, 159*, 178-187. doi:10.1016/j.apenergy.2015.08.117
34	Renn, O. (2008). Risk Participation. In O. Renn (Ed.), *Risk Governance: Coping with Uncertainty in a Complex World*. London: Earthscan.
35	Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. *Science, Technology & Human Values, 30*(2), 251-290. doi:10.1177/0162243904271724
36	Shelley-Egan, C., & Davies, S. R. (2013). Nano-industry operationalizations of "responsibility": Charting diversity in the enactment of responsibility. *Review of Policy Research, 30*(5), 588-604. doi:10.1111/ropr.12032
37	Stebbing, M. (2009). Avoiding the trust deficit: Public engagement, values, the precautionary principle and the future of nanotechnology. *Journal of Bioethical Inquiry, 6*(1), 37-48. doi:10.1007/s11673-009-9142-9
38	Stilgoe, J. (2007). *Nanodialogues: Experiments in public engagement with science*. Retrieved from London, UK: https://www.demos.co.uk/files/Nanodialogues%20-%20%20web.pdf
39	Stilgoe, J., Irwin, A., & Jones, K. (2006). *The received wisdom: Opening up expert advice*. Retrieved from London, UK: https://www.demos.co.uk/files/receivedwisdom.pdf

40	Stilgoe, J., Lock, S. J., & Wilsdon, J. (2014). Why should we promote public engagement with science? *Public Understanding of Science, 23*(1), 4-15.
41	Stirling, A. (2004). Opening up or closing down? Analysis, participation and power in the social appraisal of technology. In M. Leach, I. Scoones, & B. Wynne (Eds.), *Science and citizens: Globalization and the challenge of engagement* (pp. 218-231). London: Zed Books.
42	Sturgis, P., & Allum, N. (2004). Science in Society: Re-Evaluating the Deficit Model of Public Attitudes. *Public Understanding of Science, 13*(1), 55-74. doi:10.1177/0963662504042690
43	The Global Environmental Change Programme. (1999). *The politics of GM food*. Retrieved from Swindon:
44	Torgersen, H., & Schmidt, M. (2013). Frames and comparators: How might a debate on synthetic biology evolve? *Futures, 48*, 44-54. doi:10.1016/j.futures.2013.02.002
45	Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and socio technical change: Imagined subjectivities of 'the public' and their implications. *Environment and Planning A, 42*(4), 931-947.
46	Walls, J., Pidgeon, N., Weyman, A., & Horlick-Jones, T. (2004). Critical trust: understanding lay perceptions of health and safety risk regulation. *Health, Risk & Society, 6*(2), 133-150. doi:10.1080/1369857042000219788
47	Willis, R., & Wilsdon, J. (n.d.). *From bio to nano and beyond: A progressive agenda for technology, risk and the environment*. Retrieved from London, UK: http://www.green-alliance.org.uk/page_234.php
48	Wilsdon, J., Wynne, B., & Stilgoe, J. (2005). *The public value of science: Or how to ensure that science really matters*. Retrieved from London, UK.
49	Wynne, B. (2006). Public engagement as a means of restoring public trust in science--hitting the notes, but missing the music? *Community Genet, 9*(3), 211-220. doi:10.1159/000092659

*Reference was used in thematic analysis but is not cited in final paper.

LINKING TEXT

Chapter 3 presented several recommendations for how governments can better include publics in the context of data science. It also introduced and reviewed several areas of literature fundamental to the topic of this thesis. This included both the history of public engagement in science and technology studies as well as the current status of public engagement practice related to data technologies. The conclusions in this chapter, particularly those related to ensuring public engagement moves beyond discussion of privacy and consent as well as general information about conceptions of public engagement practice, inform the discussion and design of subsequent chapters. In particular the design of the survey in Chapter 8 uses the public engagement literature cited here to develop a scale of public engagement preferences. This chapter presents a baseline for the subsequent studies in this thesis.

From a critical perspective, this chapter is limited in its discussion on feminist justification or power concerns between publics due to its publication in a government practice journal, the audience for that journal is focussed more on practical lessons rather than theoretical discussion. Therefore the lessons described are addressing current and past public engagement practice and critiquing how this practice is problematic, but the lessons do not discuss feminism in name. However, the article and chapter are positioned to suggest change in practice and a recognition of pluralistic publics from a feminist perspective. It also offers an early point for critique on current engagement practice, which often reinforces problems of power by not including clear lines of impact on government control and regulation of data. In sum this article was both an opportunity to understand and critique public engagement practice but also an early change-making activity from a feminist perspective by publishing and engaging with a government audience.

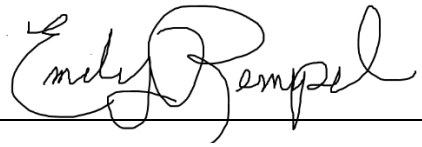
Chapter 4 builds on the results from Chapter 3 by further exploring the current practice of public engagement on government data use. Chapter 4 also address Objective 1 by aiming to examine an example of how public engagement

with data science was operationalised. This paper uses ethnography to develop lessons on the conduct of public engagement practice through the hidden curriculum perspective. This theoretical perspective, detailed in the paper, is used to analyse how underlying mechanisms influence the overt processes of public engagement. This paper further explores the recommendations and findings from Chapter 3 to critically evaluate current public engagement practice in the UK. This paper has been submitted for publication to the journal Research for All.

CHAPTER 4

The 'hidden curriculum' of public engagement with data science

CHAPTER DECLARATION

This declaration concerns the article entitled:									
The 'hidden curriculum' of public engagement with data science									
Publication status (tick one)									
draft manuscript		**Submitted**	✓	**In review**		**Accepted**		**Published**	
Publication details (reference)	Rempel, E.S., Barnett, J. and Durrant, H. (2018). The 'hidden curriculum' of public engagement with data science. *Research for All.* (submitted).								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate predominantly executed the formulation of ideas, methodology design, work, and presentation of data. Formulation of ideas: ESR primarily identified the formulation of the research question (90%) with consultation provided by JB and HD (10%). Design of methodology: ESR primarily designed the ethnographic methodology (90%) with consultation provided by JB and HD (10%). Experimental work: ESR conducted the ethnographic work, including thematic analysis of texts and notes, (85%) with thematic discussion and refining provided by JB and HD (15%). Presentation of data in journal format: ESR wrote up the thematic analysis for publication (85%) with editing provided by JB and HD (15%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							**Date**	01/06/2018	

4.1 INTRODUCTION

Public engagement with data science is an emerging field of interest following government and public attentiveness to the ways that personal and public data are used in government settings. Concerns around privacy, safe data sharing, and public awareness have entered media and government debate following events like Care.Data and Google DeepMind (Carter, Laurie, & Dixon-Woods, 2015; National Data Guardian, 2016; Powles & Hodson, 2017). Data science is the combination and application of data, including 'born digital' data like Twitter feeds and more traditional forms of 'digitized' data like administrative government records. Data science is posited to create smarter, more responsive government services (Gov.UK Blog, 2015). However, these smarter services also have the potential to do harm. For example, is it alright for children to be digitally monitored without consent to predict school and social outcomes (Vale, 2016)? Is it alright *not* to track these children if following them could flag a child in need of support? The benefit and indeed harm of these systems is dependent on individual and community perspectives related to privacy, harm prevention, political ideology, personal values, and more.

There have been a number of attempts to determine the public view on privacy issues around data science and particularly on data sharing (Cameron, Pope, Clemence, & Ipsos MORI Social Research Institute, 2014; Davidson et al., 2013; Ipsos MORI, 2006, 2016; Sciencewise, 2012, 2014). Thus far, these government public consultations have found publics hesitantly supportive of the use of personal data that is informed by a clear public benefit. However, concerns are often embedded in commercial access to personal data, and the opacity of algorithmic mechanisms like machine learning, i.e. digital processes of decision making (Cameron et al., 2014; Davidson et al., 2013; Ipsos MORI, 2016; Sciencewise, 2014). Public engagements around data use in the UK have thus far mostly been one-off workshop-style events (Aitken, de St Jorre, Pagliari, Jepson, & Cunningham-Burley, 2016) that probe public opinion on privacy and consent. There has been little critical reflection on these emerging data science

engagements, nor on how their limited focus on privacy and consent may influence what is seen as amenable to public influence. As Selin et al. (2017) argue on these fixed views of public opinion engagements, 'practices are thus often shielded from being contingent, mobile, and ultimately intertwined' (p 636). They go on to suggest that attention is often not 'paid to…the ways in which [engagement] design, publics, and findings are co-produced' (Selin et al., 2017, p. 636). Hence, the organisation of public engagement, in this case with data science, is an opportunity for critical study.

Definitions of public engagement with science and technology vary, at the most basic level it is some form of interface between individuals who develop or govern technology and 'publics'. Rowe and Frewer (2005) suggest a typology of public engagement mechanisms: communication, consultation, and participation. This categorises engagements from telling, to asking, and ultimately to including publics in science and policy. Fiorino (1990) similarly describes engagement as instrumental, substantive, or normative. Instrumental engagements include publics due to operational requirements, while substantive engagements include publics to improve a technology. Normative engagements include publics as they are due a role in technologies that may impact them. We do not offer a concrete definition of the public here but consider diverse views like Renn (2008) who categorises publics in distinct categories, i.e. as stakeholder, affected, observing or general publics versus Newman (2011) who argues that publics are called upon and produced through summoning, mediation and mobilisation by leaders.

In reflecting critically on the organisation of public engagement we draw on theories of the 'hidden curriculum' (Cribb & Bignold, 1999). The hidden curriculum, popularised in education research, exists alongside the explicit syllabus and describes imbedded methods of teaching (Cotton, Winter, & Bailey, 2013; Cribb & Bignold, 1999). For example, Cribb and Bignold (1999, p. 197) discuss the roles of 'loss of idealism' and 'emotional socialisation' in medical education. While the term hidden curriculum may suggest intentional deceit surrounding professional practices, it is better stated as a way to identify the

underlying processes related to conducting any kind of work. The 'on the surface' processes related to running a public engagement are by definition transparent, but 'below the surface' processes offer an opportunity for more nuanced critical contrast and study with these more transparent procedures. This study's aim is to examine an example of how public engagement with data science is currently operationalised from the perspective of the hidden curriculum.

4.2 METHODS AND CONTEXT

4.2.1 The ethnographic site

This study used ethnography to observe a government-led public engagement with data science in the United Kingdom. The Public Dialogue on Data Science Ethics (the Dialogue) ran from late 2015 to mid-2016 as a joint venture by the Government Data Science Partnership (the Government Digital Service, the Office for National Statistics, and the Government Office for Science), Ipsos MORI and Sciencewise. The Dialogue aimed to identify what 'the public' thinks is appropriate for government data science, inform an ethical framework, and set future goals for engagement. Eighty-eight individuals were consulted over five events in London, Sheffield, Wolverhampton, and Taunton. The workshops employed small group discussion, case study deliberation of government data science, and hypothetical deliberative scenarios. A further 2003 people were surveyed on hypothetical government data science projects to determine what was most important to the public in data science acceptability: data type, aggregate versus individual data, scope of coverage of dataset, purpose, the human role in the project, or the clarity of decisions. Results of the Dialogue are reported in 'Public dialogue on the ethics of data science in government' by Ginnis et al. (2016) and by Drew (2016), as well as an online game called 'Data Dilemmas' that allowed users to find their 'data personality'.

4.2.2 Methodology

Fieldwork over four and half months included observation and participation in planning exercises and Dialogue events. The final dataset included

ethnographic notes taken at 12 teleconference meetings, three larger face-to-face advisory group meetings, two public workshops and one government workshop, as well as the final Dialogue report and three blogs written by the project leads. The publics who participated in these events included civil servants from Cabinet Office involved in organising the Dialogue, professional public engagement organisers, other politicians and civil servants interested in data use, as well as interested members of non-governmental publics who took part in the advisory group meetings and public workshops. The majority of the ethnographic notes come from the teleconference meetings, which were attended by the public engagement organisers and the Cabinet Office Dialogue organisers. Private emails and documents were not included to protect the privacy of the organisers. Ethnography was chosen for its informal method of capturing a social phenomenon as outlined by Hammersley and Atkinson (2007). Essentially, ethnography allows for the description of the social world, in this case describing how an engagement process was undertaken. As well, Cotton et al. (2013, p. 196) argue that 'observation is widely regarded as an important tool for revealing the nuances of the hidden curriculum…since hidden curriculum research entails the search for meanings and contexts which may not be immediately visible to actors in that context.' Inductive thematic analysis was used to analyse the note and text data, using an iterative process of coding, theme identification, and review (Braun & Clarke, 2006) using NVivo 10. Initial codes were reviewed by all authors and a final code set was developed and subsequently used to code the remaining notes and text documents. The final codes were then again discussed by all authors to group in to themes. While the themes are set within the context of data science, they may also be relevant to other technologies that are subject to government-led public engagement. This study received ethical approval from the Department of Psychology's Research Ethics Board at the University of Bath.

4.3 Results and discussion

There were eight lessons or themes identified through the ethnographic study. We will discuss those most relevant to the 'hidden curriculum', e.g. those things

that shaped on the surface processes, namely: how organisers identified a public, setting the purpose of the dialogue, the variations in nomenclature, and how interests of stakeholders were negotiated, particularly through the theme of setting the discussion parameters. We end with a brief discussion on the remaining three: time as a key constraint, how meaning was interpreted from the dialogue materials, and mechanisms of validation through publication.

4.3.1 Lesson 1: Defining the public/Imaginations are made reality

Inherent to any public engagement, and indeed this Dialogue, is *identifying and imagining a public*. While there was relatively unproblematic referral to the Dialogue participants as 'the public', organisers and advisory group members made several references to seeking opinions from the 'average' person, whom was imagined to be neither highly critical nor extremely supportive of data science. Members of this group were not imagined to be activists nor highly literate in data science methodologies. Data science literacy was seen as a key group identification benchmark, e.g. workshops were held with 'techy' and 'high data' user groups in order to separate the general and specialist publics. In writing up the results, efforts were made to report a diversity of opinions from these different groups. This was seen to demonstrate to external parties that participants were part of the ideal public, i.e. neither too positive nor too critical. In fact, a chapter of the final report was devoted to describing differing perspectives to highlight that the public held a mix of notionally positive opinions on data science. From the hidden curriculum perspective, we can see these imaginations of the 'average' opinion and separating out the groups to different levels of data literacy as multifaceted practises that shape who the public are. Barnett, Burningham, Walker, and Cass (2010) and Walker, Cass, Burningham, and Barnett (2010) argue that political actors hold pre-conceived conceptions or imaginations of the public and what they believe. This theme demonstrates how pre-conceptions can be used to define and shape the kinds of publics that are enrolled in engagement exercises. By imagining the 'real' public as being neither overly critical nor supportive, these sorts of individuals are then targeted for enrolment in the engagement exercise. While this

reflects theoretical literature around the fluid and politically constructed nature of the public (Newman, 2011) it also demonstrates how the theoretical process of imagining the public lends itself to limiting who is and is not the 'public' in public engagement (Mahony & Stephenson, 2017). The imaginations are made reality.

4.3.2 Lesson 2: Setting the purpose of the Dialogue/Developing explicit and meta-objectives

The dialogue reported several explicit objectives as outlined by Ginnis et al. (2016) that have the potential to align with substantive engagement, for example having the public comment on the ethical guide (Fiorino, 1990). If the motivation behind this is to improve the guide it fits within a substantive aim, while if the motivation is to avoid future public opposition then it aligns more closely with instrumental engagement. These motivations can be explored more clearly through the process of developing underlying or meta-objectives. One key example was the aim to anticipate and assuage public concerns around data science. The Dialogue process was seen as a way to explore and counteract potential public fears and therefore build a positive space for government data science. As MP Matt Hancock stated the guideline's purpose was to help "people in government to feel confident using new techniques" (Hancock, 2016). In fact, during the planning sessions there were several reflections on how this dialogue may help avoid highly publicised negative data events. These underlying purposes to 'counteract fears' were not unproblematic to the organisers. Some raised concerns about this instrumental objective of communicating and calming. And while there were efforts to ensure the dialogue allowed for open communication from publics, these hidden objectives persisted. As Felt and Fochler (2010, p. 228) argue engagement can be a way "surveying and assessing potential critical voices" and this was evident in this Dialogue. The Dialogue's more explicit aims of public empowerment seem at odds with the discourse and planning around empowering data users. It is difficult to determine in final reports where the meta- or explicit objectives drove the conclusions. This process

of holding two opposing kinds of objectives may suggest that public engagement practice despite being explicitly substantive may be implicitly instrumental.

4.3.3 *Lesson 3: Engaging engagement/The meaning of engagement is contested*

Organisers, advisory group members, and publics held varying conceptions of what was meant by *engagement*. At times engagement represented public inclusion, particularly when discussing the workshop components, but was also referred to as generating interest, e.g. the online tool should be engaging for users. This led to confusion during advisory group meetings, which included a wide group of stakeholders, where there were evident contestations as to what engagement as an action comprised. For some, engagement was public participation while for others engagement was seen as education. These contested meanings were exemplified by discussion of the online 'data dilemmas' game. It was often referred to as an engagement tool, but also as an education or knowledge tool. The final report stated that the 'engagement tool [will be used] to engage a wider audience in a public debate around data science' (Ginnis et al., 2016, p. 10). One can see how contested meanings could alter the substance of this sentence. If engagement is evoking interest and the tool excites publics, then the work is done. If engagement is participation the tool would need to be followed by more active public involvement. The former meaning suggests participating in the game is engagement and the latter suggests participating in the game should enable further engagement. While authors like Smallman (2016) demonstrate that academic researchers have in recent years intended engagement to be participative, other meanings are more common outside of academia. The term engagement is multi-dimensional. In practice, one cannot assume that the academic shorthand of engagement as public participation is consistent in government or industry settings where engagement more often means communication. Establishing a common language between academia, publics, government, and industry may help all parties critically reflect on the meaning of engagement and thus how to encourage more inclusive and substantive engagement practice.

4.3.4 *Lesson 4: Negotiating stakeholder interests*

Throughout the Dialogue, the organisers acted as intermediaries between the wider stakeholder group consisting of public engagement and data science professionals, interested government departments, and the participants. They negotiated expectations, particularly those of the government departments, through anticipating what could be sensitive topics. For example, discussion around terrorism was considered to be a sensitive case. There was ongoing adaptation to these anticipated problems to ensure that they did not become roadblocks. One of the principal concerns was balancing the tone of the engagement, i.e. ensuring the report did not read either too positively or too negatively. It was anticipated that participants would want a representative account of their discussions, while government would want a positive tone, and a negative tone at times would be necessary to balance out the arguments. Organisers had to develop, whether real or perceived, knowledge and imagination of 'hot topics' for both government and publics. The dialogue materials, the final report, the case studies of government data science, and more can be seen as objects processed to be as uncontested as possible. For example, a sub-theme included deciding what sort of materials to present to publics during the engagement. This process of *setting the discussion parameters* is also a way of drawing lines around what can be said. The Dialogue used real government case studies to stimulate debate, which were designed to test certain perceptions of public concern, e.g. keeping data secure. Thus, running the workshops was not simply a process of organising materials and resources but was also a constant negotiation of what could be said and what should be said both to publics and to stakeholders. This anticipation of concerns furthers the understanding of public engagement as a process of preventing critique (Felt & Fochler, 2010). This theme suggests that further than public engagement identifying critique, it is also a practice of anticipating concerns for the engagement itself.

4.3.5 Other lessons and practices of interest

There were several underlying practical constraints during the Dialogue, including *managing time and public understanding*. From inception meeting to final report, the Dialogue lasted just six months. This short timescale left little time for reflection during the Dialogue process and organisers were in a constant cycle of iterating plans, reports, and workshop materials. At one point during the project the team piloted an education session on a Friday night and proceeded to re-configure it for another session on Saturday morning. While organisers planned to allow spontaneous suggestions for ethical concerns, there was very limited time after the education and deliberation components for this process. The discussion was then limited to these pre-determined ethical concerns. These limitations are then, necessarily, reflected in the reporting where *making meaning is performed.* As Lezaun and Soneryd (2007, p. 288) argue the reporting stage is about "attempting to officialise a singular meaning for the exercise." In order to maintain clarity of meaning, the report was written with different prospective readers in mind, e.g. media versus government, and organisers focused on avoiding writing in a way that could be interpreted negatively. The final, and perhaps most practical, step in the engagement process was releasing these results. A final stakeholder event was held to discuss the report with a wider group of interested parties. In interacting with these larger groups, there was a focus on not overstating the applicability of the results. For example, the final report states that "the views of proportions of the qualitative group should not be extrapolated to the population at large" and that "[the] results are intended to be illustrative rather than statistically reliable" (Ginnis et al., 2016, p. 66). These caveats can function as a way of safeguarding the organisers from these aforementioned negative meanings.

4.3.6 Reflecting on the author's role in the engagement

My role in the Dialogue was as an observer, and at rare times a participant in larger meetings. However, due to my limited role there were meetings and communications that I was unable to observe. It's impossible to know if following more private exchanges would have changed the nature of the themes discussed here. The majority of my notes come from teleconference meetings which also risk

missing non-verbal cues. However, this allowed me to take notes freely without attracting notice. I found it challenging to feel comfortable in my role, being a non-expert in the design of public engagement exercises, but being unaware of the status quo of public engagement allowed me to highlight processes that I may not have recognised as hidden knowledge had I been more familiar with government-run public engagement. As time went on I felt more comfortable as an observer and found ways to give myself a sense of value in the process. Although I attempted to minimise incorporating my own views on data science and public engagement in my interpretation of the process, as I became more familiar with the organisers I expect my comfort led to unconscious self-filtering to avoid reflecting the Dialogue in a negative light. Evaluation of any dialogue is challenging, as Rowe (2005) highlights, it is difficult to define effectiveness in public engagement. As such, and due to the limitations of ethnography, this study is meant to be descriptive rather than evaluative.

4.6.7 Critical reflection

These themes exist within established practices of public engagement professionals, e.g they were conducted by Ipsos Mori who commonly run and organise these kinds of public engagements. Due to this context the lessons found here may be familiar to those who design or have taken part in 'professional' engagements at the national level, however it is important to take note of this context when considering public engagement at large. Public engagement in different contexts, for example at the local level, may not experience these same kinds of hidden norms. For example, in local government interactions with publics may often be directly between civil servants and non-governmental publics themselves and not mediated by paid public engagement professionals. Public engagement in other contexts likely does not have the same financial and personal resources to draw on in designing and running public engagement. Further research and reflection is needed on what established public engagement practice with data technologies looks like in these different government contexts.

4.5 CONCLUSION

Public engagement with data science is becoming increasingly popular in an attempt to understand where the social and ethical lines should lie in government data science. The Public Dialogue on Data Science Ethics is one such example. This ethnographic study examined this Dialogue to understand how a public engagement with data science is operationalised from the perspective of the hidden curriculum. Embedded within practical considerations of organising meetings and creating stimulus plans is the underlying processes of negotiating stakeholder interests, identifying who the public is, and setting the purpose of the Dialogue. There was also a key lesson in defining what engagement consists of, examples in this Dialogue included an 'engaging' online quiz-style tool and workshops that were a public 'engagement'. With cautious interpretation that this Dialogue's processes are in some manner reflective of wider trends, the authors conclude by questioning how these processes of doing engagement may impact the space for public involvement in data science now and in the future.

The hidden curriculum perspective suggests that despite the explicit descriptions in official documents and reports that public engagement is about substantive and inclusive participation of publics, in practice the norms of working still align more closely with an instrumental, top-down approach to public participation. While there is no clear solution to how to address the conflict between the hidden curriculums of public engagement and how it is presented in official reports, a more transparent practice of releasing the materials and content used during public engagement exercises may be of use. In addition, practitioners and academics should think critically about what they mean by terms like engagement, as well as how the sorts of compromises made during the conduct of engagement like negotiating conflict may impact the results of that engagement. With an eye on brevity, we also suggest looking to data science itself for potentially innovative public engagement opportunities. For example, civic hacking in government and crowd-sourcing of data collection and algorithm development (Lauriault & Mooney, 2014; Schrock, 2016; Sousa, 2013). We end by

also suggesting further critical reflection on the hidden curriculum in public engagement processes and how these normalised ways of 'doing' engagement can expand or shut-down public participation in science and technology.

REFERENCES

Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016). Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics, 17*(1), 73. doi:10.1186/s12910-016-0153-x

Barnett, J., Burningham, K., Walker, G., & Cass, N. (2010). Imagined publics and engagement around renewable energy technologies in the UK. *Public Understanding of Science, 21*(1), 36-50. doi:10.1177/0963662510365663

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Cameron, D., Pope, S., Clemence, M., & Ipsos MORI Social Research Institute. (2014). *Dialogue on Data - Exploring the public's views on using linked administrative data for research purposes*. Retrieved from London, UK: https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics, 41*(5), 404-409. doi:10.1136/medethics-2014-102374

Cotton, D., Winter, J., & Bailey, I. (2013). Researching the hidden curriculum: intentional and unintended messages. *Journal of Geography in Higher Education, 37*(2), 192-203. doi:10.1080/03098265.2012.733684

Cribb, A., & Bignold, S. (1999). Towards the reflexive medical school: The hidden curriculum and medical education research. *Studies in Higher Education, 24*(2), 195-209.

Davidson, S., McLean, C., Treanor, S., Ipsos MORI, Aitken, M., Cunningham-Burley, S., . . . University of Edinburgh. (2013). *Public acceptability of data sharing between the public, private and third sectors for research purposes*. Retrieved from

Drew, C. (2016). Data science ethics in government. *Philsophical transactions of The Royal Society A - Mathematical physical and engineering sciences, 374*(2086).

Felt, U., & Fochler, M. (2010). Machineries for Making Publics: Inscribing and De-scribing Publics in Public Engagement. *Minerva, 48*(3), 219-238. doi:10.1007/s11024-010-9155-x

Fiorino, D. J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. *Science Technology Human Values, 15*(2), 226-243.

Ginnis, S., Evans, H., Boal, N., Davies, E., Palmqvist Aslaksen, A., Ipsos MORI Social Research Institute, . . . Sciencewise. (2016). *Public dialogue on the ethics of data science in government*. Retrieved from London, UK: https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf

Gov.UK Blog. (2015). Data Science ethics. Retrieved from https://data.blog.gov.uk/2015/12/08/data-science-ethics/

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in Practice* (Third ed.). New York: Routledge.

Hancock, M. (2016). *Data science ethical framework launch: Matt Hancock speech*. London: Cabinet Office Retrieved from https://www.gov.uk/government/speeches/data-science-ethical-framework-launch-matt-hancock-speech.

Ipsos MORI. (2006). *The use of personal health information in medical research general public consultation: Final report*. Retrieved from London, UK: https://mrc.ukri.org/documents/pdf/the-use-of-personal-health-information-in-medical-research-june-2007/

Ipsos MORI. (2016). *The one-way mirror: Public attitudes to commercial access to health data*. Retrieved from https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Lauriault, T. P., & Mooney, P. (2014). *Crowdsourcing: A geographic approach to public engagement*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2518233

Lezaun, J., & Soneryd, L. (2007). Consulting citizens: technologies of elicitation and the mobility of publics. *Public Understanding of Science, 16*(3), 279-297. doi:10.1177/0963662507079371

Mahony, N., & Stephenson, H. C. (2017). Engaging with the public in public engagement with research. *Research for All, 1*, 35-51.

National Data Guardian. (2016). *National data guardian for health and care review of data security, consent and opt-outs*. Retrieved from London, UK: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF

Newman, J. (2011). Public leadership as public-making. *Public Money & Management, 31*(5), 315-322. doi:10.1080/09540962.2011.598336

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*. doi:10.1007/s12553-017-0179-1

Renn, O. (2008). Risk Participation. In O. Renn (Ed.), *Risk Governance: Coping with Uncertainty in a Complex World*. London: Earthscan.

Rowe, G. (2005). Difficulties in evaluating public engagement initiatives: reflections on an evaluation of the UK GM Nation? public debate about transgenic crops. *Public Understanding of Science, 14*(4), 331-352. doi:10.1177/0963662505056611

Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. *Science, Technology & Human Values, 30*(2), 251-290. doi:10.1177/0162243904271724

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society, 18*(4), 581-599. doi:10.1177/1461444816629469

Sciencewise. (2012). Public dialogue on data openness, data re-use and data management: Sciencewise co-funded project 2012. Retrieved from http://www.sciencewise-erc.org.uk/cms/public-dialogue-on-data-openness-data-re-use-and-data-management/

Sciencewise. (2014). Big Data: Public views on the collection, sharing and use of personal data by government and companies.

Selin, C., Rawlings, K. C., de Ridder-Vignone, K., Sadowski, J., Altamirano Allende, C., Gano, G., . . . Guston, D. H. (2017). Experiments in engagement: Designing public engagement with science and technology for capacity building. *Public Underst Sci, 26*(6), 634-649. doi:10.1177/0963662515620970

Smallman, M. (2016). Public understanding of science in turbulent times III: Deficit to dialogue, champions to critics. *Public Understanding of Science, 25*(2), 186-197.

Sousa, S. (2013). *Lessons learnt from a public–private big data hackathon*. Retrieved from London, UK: https://fvstatic.s3.amazonaws.com/1425646063_0599941001425646063.pdf

Vale, R. (2016). The register of schools in England goes into alpha. Retrieved from https://data.blog.gov.uk/2016/11/29/the-register-of-schools-in-england-goes-into-alpha/

Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and socio technical change: Imagined subjectivities of 'the public' and their implications. *Environment and Planning A, 42*(4), 931-947.

LINKING TEXT

Chapter 4 identified several underlying processes of public engagement on government data science that influence the overt results of public engagement. These include stages of finding an ideal public, using underlying meta-objectives, negotiating conflicting interests from stakeholders, as well as other processes including confusion on the meaning of the word engagement. These themes, alongside the findings in Chapter 3, suggest that public engagement is often not reflective of broader aims like decentring the responsibility for data governance nor of challenging power relations. Thus, the baseline for public engagement with government data usage that is built on in subsequent chapters is one in which publics are not often meaningfully included in decisions made around data technologies. Public engagement is still often a practice of one-off, small-scale and distributed events that are not necessarily reflective of the pluralistic and power-challenging aims of decentred governance and feminism. From a feminist perspective, these themes help to establish where opportunities for change exist in current practice of calming and quieting publics' contestations. For example, recognising and counteracting meta-objectives that can undercut the substantive purpose of public engagement. It also highlights how publics were limited to small-group discussions and survey taking and were not open to be involved in the overall public engagement planning or design. Publics were limited in their scope for influence on both the engagement process itself but also the ultimate data science ethical guide. This highlights the way that power is still firmly held within government publics and even within public engagement professionals and that these exercises highlight the lack of power that non-governmental publics hold. Public engagement practice should better consider how to avoid re-enforcing problematic power relations. This chapter concludes the thesis' exploration of current and past practice in public engagement within the UK.

Subsequent chapters move on to explore the context for engagement (i.e. how governments use data) and the future of engagement (i.e. how publics want to and could be engaged). Chapter 5 examines the former in relation to Objective

2. This chapter along with Chapter 6 present sister papers both seeking to examine the nature of government data use. In addition to exploring data use, this chapter discusses and proposes the story completion method for use in policy research. Although this method is relatively well-established in social psychology research, it is not widely known in other disciplines. In this study, local government workers and data analysts were asked to complete a short story related to data sharing within their local area. This chapter aims to first explore how local government workers imagine hypothetical data projects. Second to explore how this may or may not reflect the contingencies of data projects in local government. And third to examine whether story completion is a viable method for understanding policy and data processes. The purpose of the following two papers, from a feminist perspective, is to understand the landscape of data processes in order to understand how change can be enacted in public engagement practice. This paper is not currently submitted for publication.

CHAPTER 5

Story completion as an exploratory method into the rhetoric around data use in local UK government

CHAPTER DECLARATION

This declaration concerns the article entitled:									
Story completion as an exploratory method into the rhetoric around data use in local UK government									
Publication status (tick one)									
draft manuscript	✓	**Submitted**		**In review**		**Accepted**		**Published**	
Publication details (reference)	N/A								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate predominantly executed the formulation of ideas, methodology design, work, and presentation of data. Formulation of ideas: ESR formulated the research idea and plan (90%) under the direction of JB and HD (10%). Design of methodology: ESR designed the story completion exercise and research design (85%) with consultation from JB and HD (10%) as well as review by a third party (5%). Experimental work: ESR conducted the research and thematically analysed the stories (90%) with editing assistance from JB and HD (10%). Presentation of data in journal format: ESR wrote up the results in journal format (85%) with editing and supervision from JB and HD (15%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						**Date**	01/06/2018		

5.1 INTRODUCTION

Secondary data use, big data, data science, and data systems are increasingly promoted by government in the UK to improve government services and cut costs. In 2017, the United Kingdom's Chief Executive of the Civil Service stated that "[d]ata is at the heart of 21st century government…It makes government work for everyone, by better reflecting the world that we live in" (Manzoni, 2017). In 2016 the then Minister for the Cabinet Office claimed that "[i]t is vital we seize the opportunities that data science presents. The biggest risk would be to do nothing" (Hancock, 2016). This reflects the claim that "[d]ata is increasingly becoming a source of wealth and public value creation. In that context, one can argue it is more valuable than just being "the new oil". It is the lifeline of the digital society" (Jarrar, 2017). There is an ever-present framing of both local and national government data use as key to the future of civic society. With such a strong narrative around the inevitable good of data science, it is challenging to unpick rhetoric from reality. And while some authors offer critical appraisals of government readiness for this future (Milakovich, 2013), few explore the social, political, and technological realities of government data use. Without critical reflection on these realities, it will be challenging to build data systems that 'work for everyone'. In particular, research is needed in local contexts where the drive to innovate with data is underpinned by scarce digital and human resources (Cornford, Wilson, Baines, & Richardson, 2013; Malomo & Sena, 2017). In this paper we discuss how novel qualitative methodologies can be used to explore the contingencies of data science in local UK government. We begin by setting definitional clarity around data science in local government. We then discuss the use of qualitative narrative methods in policy research and introduce the story completion method. We discuss how to design case studies using story completion and follow this by presenting the results of an exploratory pilot study in a local authority in the South West of England. We then discuss how these results relate to the rhetoric and reality of data sharing in local government. In conclusion we

seek to characterise the validity of this approach, as well as consider how story completion could be used in wider policy contexts.

5.1.1 Data in (local) government

Governments have always used data, albeit with different forms, methods, and intentions. Since the mid-2000s terms like 'big data' and 'data science' have slowly infiltrated government rhetoric and practice. A search on Gov.UK (Government Digital Service, 2018), the National UK government's online portal, shows 138 of 644 speeches in 2017 contained the word 'data'. In other words, one in five speeches used 'data' as a talking point. From a technical stance, big data can be defined as data collected automatically and quickly at massive scale (Ward & Barker, 2013). Examples include web traffic and mobile GPS data. Whereas data science is defined as the procedures related to combining and creating with data. However, big data and data science can also be defined for their social dimensions, like Markus and Topi (2015, p. 3) who call for "a sociotechnical perspective, viewing it as a cluster or assemblage of data-related ideas, resources, and practices." Such a perspective allows greater consideration of the increasing role of data in governance and day-to-day life. Popular terms include 'the datafied society', 'data communities', 'digital governance', 'data infrastructures', and 'datafication' (Data Power, 2017). These terms, though imperfect, more fully capture the ways that as individuals we interact with and are influenced by data. Everything from shop purchases to online behaviour to mortgage allocation is tracked, monitored, and influenced by data and related data technologies. Milakovich (2013, p. 1) suggests government could use data for everything from terrorism to healthcare in order "to find new ways to analyse metrics and maximize internal operational efficiencies". The main motivation being that data analytics has the potential to both improve service provision and save costs. In this context we propose the term 'civic datafication' which we intend to represent the ways in which data and data systems are increasingly inscribed not only in social interactions but in the ways that citizens interact with government from local to international contexts. Civic datafication recognises the social and political

dimensions of big and new forms of data. Civic datafication is therefore not only about how governments increasingly use data and data technologies, but also how the community and the citizen is fundamentally located in the use and creation of these technologies.

The model of civic datafication is increasingly common within local government in the UK (Malomo & Sena, 2017). Several studies have described pilots of local data projects. For example, Taylor et al. (2015) used participatory data design encouraging local residents of a single street in the city of Cambridge to work on data projects about their local area. The smart city movement has also been promoted as a way to encourage the uptake of digital technologies, data, and civic participation at the local level. It is described enthusiastically by Keller, Koonin, and Shipp (2012, p. 5) as "knowing practically every detail about a city. The state of its infrastructure, its inhabitants, its environment are all known to you, to high resolutions in time and in space." Despite this positive rhetoric around civic datafication, the implementation is limited. Several authors (see: Cornford et al., 2013; Keller et al., 2012; Kitchin, 2013) have highlighted "that for a long time the ambition around the development of Big Data capabilities has not matched the actual use of analytics in local governments" (Malomo & Sena, 2017, p. 8). And despite these critiques there has been little to no empirical investigation of how and why local governments are failing or succeeding to take up the model of civic datafication. A challenge of working in a setting like local government where resources are scarce and time is limited (Keller et al., 2012), is that such reporting is rare. Furthermore reporting 'failures' in data projects may be even less likely as it is incumbent upon local government to be doing well for their communities and to show that they are doing well. How then do we incorporate elements of criticality and reflexivity in local government datafication? How do we discuss both failures and successes? In the following sections we present the story completion method as one potential option to answer these questions.

5.2 STORY COMPLETION AS A METHODOLOGY

5.2.1 What is story completion?

Story completion is a prompt-based qualitative method traditionally used in both psychology research and psychological treatment (Clarke, Hayfield, Moller, & Tischner, 2017). It is a narrative, mostly text-based completion task similar in nature to a structured interview question but includes a more creative, in-depth prompt. Although modern story completion has its roots in work done by Kitzinger and Powell (1995) in the mid-1990s, more recently it has been popularised by Clarke et al. (2015) and Clarke et al. (2017) within studies on sexuality, infidelity, and body hair modification. It is used to study sensitive topics like these as it explores hypothetical and abstract ideas that may be challenging to discuss directly. It is, therefore, a projective technique that aims to explore perceptions rather than personal experiences. For example, eliciting the responses that young adults imagine would ensue from revealing their homosexuality to their parents. Story completion studies are structured around three stages. First the design of the story prompt which can be first or third person and may include comparative prompts for experimental hypothesis testing. Second, participants complete the prompt verbally, on paper, or digitally and third, the final stories are collated and analysed in a range of ways, most commonly through thematic, discourse, or narrative analysis. The choice of analysis will depend on the researcher's own epistemological perspective. Long-standing versions of story prompts like the Attachment Story Completion Task (Bretherton, Ridgeway, & Cassidy, 1990) have employed quantitative analytical techniques through established coding frameworks (Mikic & Terradas, 2017; Ştefan & Avram, 2017). Clarke et al. (2017) argue that story completion can be used within both a social constructionist and a post-positivist epistemological stance. However, as story completion is a projective technique, we perceive its core strength to be enabling us to reveal how data projects are discursively constructed by those that are engaged in them, and what this tells us about socio-cultural practices of data use at a local government level. Therefore, we approach story completion for civic datafication from a social constructionist perspective. For further discussion on the development of story completion and its' historical uses see Clarke et al. (2017).

Box 5.1: Example Story Completion Stems

From Moore (1995)	**From Clarke, Braun, and Wooles (2015):**
(1) Jenny rang Sue to ask her to go swimming. She has her period, what does she do? *(2) Mary tells her mother that she has her period for the first time. Her father comes into the room….What will happen now?* *(3) Jill has cramps in her tummy which she gets on the first day of her period. What does she do?* *(4) Ben loses a book belonging to Anne. Anne is cross with him. Ben says "You are in a bad mood because you've got your period?" What does Anne do?* *(5) Nicole was standing at the bus stop when she noticed a couple of boys who seemed to be laughing and talking about her. She turned around to pick up her schoolbag and noticed a big reddish-brown stain on the back of her dress. What do you think happened next? How did it end?*	*Sarah wakes up early on Tuesday morning and follows her usual routine of getting ready for work while John, her husband of four years remains sleeping. Later that day Sarah returns home early from work, as she enters the house she notices John's coat and work shoes in the hall way. Thinking he must have come home from work sick she walks upstairs to their bedrooms, when she opens the door she is confronted with John in bed with another man...* *Sarah wakes up early on Tuesday morning and follows her usual routine of getting ready for work while John, her husband of four years remains sleeping. On her lunch break Sarah decided to try out a new café that a work colleague has recommended. As she walks towards the café, much to her surprise she notices John sitting at one of the tables outside with a man she has never seen before. As she gets closer she notices that John is holding hands with the man and he is smiling and gazing into the man's eyes…*

5.2.2 Designing a story completion study

Story prompts can be as detailed or vague as required. A more detailed prompt will invite a lower range of responses, while a vague prompt may be challenging for participants to complete. Clarke et al. (2017) discuss a range of factors to consider in prompt design including length, authenticity, how engaging the prompt is, the amount of detail, the deliberate use of ambiguity, and first versus third person narratives. In short, a good prompt should be detailed enough that participants feel confident and interested to answer it. In Box 5.1 we present several examples of story completion prompts (Clarke et al., 2015; Moore, 1995). Clarke et al. (2015) employ comparative prompts to explore perceptions of emotional versus sexual same-sex infidelity which may be difficult to uncover in

an interview setting, particularly as participants may not have personally experienced same-sex infidelity. Moore (1995) also utilises comparative prompts for a study about adolescent Australian girl's perceptions of menarche, these prompts highlight the need to consider the age and background of the prompt respondents, e.g. young girls versus adults. As other authors have highlighted, prompts should always be piloted to ensure they convey what the researcher intends.

Other aspects of story completion studies to consider include sample size, completion instructions, and mode of completion. Online completion has the advantage of easy collation and speed (Braun & Clarke, 2013; Clarke et al., 2017) but completion modes vary, for example verbal response with children using dolls to tell a story (Bretherton et al., 1990; Ştefan & Avram, 2017). Completion instructions also vary and can specify time taken or length of story. Some authors specify respondents take at least ten minutes to complete their stories (Clarke et al., 2015) while others simply ask participants to finish the story (Ştefan & Avram, 2017). Ideally instructions should ensure stories are long enough to produce detailed and meaningful responses. Finally, researchers must consider sample size. A larger sample will be needed for quantitative analysis, while the sample size for qualitative analysis will depend on the research question. Braun and Clarke (2013) suggest, alongside critical reflection on your research question, that 15 to 40 stories are sufficient for a small study while 200 plus is more appropriate for larger studies. However, identifying appropriate sample size is challenging in qualitative research and some authors caution against using 'rule of thumb' guidelines (Malterud, Siersma, & Guassora, 2016). Sample size sufficiency instead should be thought of as contextualised within the study itself related to as Vasileiou, Barnett, Thorpe, and Young (2018) discuss data adequacy. While it is tempting to say more is better, it is better argued in story completion that the sample size should provide adequate evidence, both in variety of stories and consistency, to support the research question.

5.2.3 The boundaries and benefits of story completion for civic datafication

Story completion has yet to expand meaningfully outside of psychology. A search dated February 4th, 2018, for "story completion" in either the title, abstract, or keywords resulted in 312 results in PsycINFO, 217 in Scopus, and 128 in Web of Science. The vast majority of results, therefore, coming out of the Psychology-themed search engine PsycINFO. While civic datafication may not seem at first glance a logical fit to story completion, we propose that it is an ideal technique for four main reasons:

1. First that story completion allows for projective exploration. Studying rhetoric is in essence studying perceptions of data in government. Story completion is one route into characterising and understanding social discourses about data use.
2. Second, story completion allows for broad and creative responses. It encourages exploring civic datafication from a qualitative, social constructionist epistemology in an area of research that has been dominated by quantitative and post-positivist perspectives.
3. Third, from our own previous research as well as conversations with government researchers, we see the use of data as being organisationally sensitive. Individuals may be hesitant to share cases where data projects did not succeed or exhibited unethical data practices like data loss. Story completion allows researchers to examine the contingencies of data projects through hypotheticals.
4. Fourth, story completion complements a history of both narrative analysis of policy-making processes and understanding policy as a narrative within policy-making. Policy-making is often about social discourse and narrative-making (Jones, Shanahan, & McBeth, 2014; Kettl, 2016), story completion allows for a novel method for exploring how social discourse shapes policy and vice versa.

We turn next to discuss a pilot story completion study conducted in local government in the UK.

5.3 PILOT STUDY

5.3.1 Pilot study methodology

We conducted pilot study in a local authority in the South West of England. The aim of this project was three-fold. First to explore how participants imagine hypothetical data projects. Second to explore how this may or may not reflect the contingencies of data projects in local government. And third to examine whether story completion is a viable method for understanding policy and data processes. We identified two local data projects through an established relationship with a manager within the local authority. The first was an internal project within the local authority and the other was a collaborative project between community volunteers and the local authority. Participants were recruited through a snowballing method where each participant was asked if they knew other individuals who worked on the data project in order to include all individuals who were involved in any aspect of the data project. In total 15 individuals were suggested, of which we contacted 13 as two were later identified to not have worked on the project directly. Two individuals declined to participate, and one did not respond to our request. The final sample consisted of 10 individuals who responded to a short prompt about data sharing in their local area. All participants were either local government workers or community data volunteers who worked on the two data projects. These participants were part of government and technical publics who specialise in the use, regulation, and analysis of data. They are mostly civil servants including policy managers, data analysts, and administrators, as well as technical specialists who work alongside government as civic analysts. Therefore the participants do not belong to policy-making publics and are not politicians, their responses and discussion to the story prompt should be understood within these parameters.

Participants were given the option to take as little or much time as they required to fill in the story by hand. The task was conducted immediately following traditional semi-structured interviews that asked questions about the facilitators and barriers to data projects. The prompt stated “Sam is a local

community organisation manager. She/He wants to understand how to get access to local council data about the people who use her/his organisation's services. She/He heard that there were local people in the council who she/he could contact. What does Sam do next?" A short paragraph of instruction was presented prior to the prompt which stated "Please read the following prompt carefully and complete the story. It can be as long or short as you like, feel free to take extra space on the back of this form if you need it. There are no wrong answers. Thank you!" The prompt was reviewed for coherence between the authors as well as with the key contact at the local authority. We chose a prompt about data sharing to ensure it could be answered by a range of respondents, regardless of their technological expertise. Data sharing between a community organiser and local government also allowed for stories related to the broader social dimensions of civic datafication as it did not relate solely to technological process. After data collection, the prompts were collated and thematically analysed using NVivo 10. The thematic coding processes followed the Braun and Clarke (2006) methodology of first reviewing, identifying potential themes, then re-reviewing and grouping into common themes.

5.3.2 Pilot study results

The average story length was 114 words while story structure varied widely. Several participants wrote in a numbered sequence or bullet point, while others wrote paragraph style. Most participants were confused at first on what they should write and were encouraged to write whatever they felt fit the story. Some participants positioned themselves as the data provider in the story, providing advice to Sam on how to get data from themselves while most simply wrote pieces of advice for Sam for gaining access to data in general. Most participants did not expand on Sam as a person or the kind of organisation she ran, however a few assumed that she ran a commissioned local service. Overall, participants seemed to engage with the task and while some stories were much shorter than ideal, all stories were completed. From the ten stories, five major candidate themes were identified alongside one minor theme. We consider these

as candidate themes as the exercise was a pilot and we reflect on how this work can best be taken forward in the discussion section. The five major themes include (1) the importance of a personal connection, (2) the need to align organisational and individual priorities, (3) considering alternative ways of accessing data, (4) ensuring ethical access and data handling, and (5) describing and highlighting the complexities of data flows. The one minor theme is (5a) questioning the value and likelihood of an outcome. Story quotes within each theme are presented as written and were not edited for grammar or punctuation.

Theme 1: Personal connection

A key feature in the stories that followed logically from the prompt was the need for a personal connection in order for Sam to gain access to the data. This was alluded to in a range of ways; from broad sentiments like "*Write/contact the relevant people"* [Participant 4] to mentions of specific organisations or individuals who should be contacted within the local authority. Underlying the generalised need to contact the right people, was the concept of getting "*sympathetic people"* [Participant 2] on side. This was sometimes characterised as developing relationships wherein Sam would be more likely to succeed if she was able to convince others to help her. Having a personal connection and knowing the right people to contact was seen as essential for Sam. Sam, therefore, needed to build working knowledge of who was the right person to contact. Stories often featured a chain of such contacts, for example "*Sam needs to get in touch with the council. On doing this the contact at the council should be able to point Sam in the direction of the relevant directorate that Sam needs to speak to*" [Participant 8] or "*she may well have a few conversations with people in the [organisation] before ending up with me*" [Participant 10]. Sometimes this initial communication was predicated on Sam submitting a Freedom of Information (FOI) request or submitting other kinds of formal requests for data. However, all stories relied on Sam succeeding in finding the right people to help her.

Theme 2: Organisational versus individual priorities

Closely related to finding sympathetic people within the local authority, was Sam ensuring her reasons and aim for accessing this data aligned with the council's priorities. This was characterised in one story as finding "*a common vision + agreed set of outcomes*" [Participant 4]. And in another as finding someone "*who can champion the idea of open data*." [Participant 2]. Several stories highlighted that Sam's success would be dependent on why she wanted data and her ability to describe the benefit of her receiving it. For example, access to it was contingent on describing "*what outcomes/benefits could be gained for sharing this data*." [Participant 4]. While it is reasonable to expect that the council would not release data to someone without proper justification, it is important to note how access to data was seen as politically motivated. Access depended on Sam's justifications matching political goals within the council, for example the aforementioned "*idea of open data*". Participants did not comment on why they thought Sam wanted the data but rather suggested, regardless of the reason, she should tailor her justification to show benefit to the council. Data access was therefore contingent on being mutually beneficial to both Sam and the council.

Theme 3: Alternative data

While the prompt specified that Sam had heard there were local people she could contact to access data about her own organisation, several stories rejected her accessing data directly through this contact. As previously mentioned, some participants suggested she could fill in an FOI request while others suggested she look elsewhere for data. Some stories also mentioned that "*Sam should already know a level of info on data*" [Participant 7] if she ran a commissioned service. A few participants anticipated Sam having trouble accessing data through the council and suggested alternative data sources, for example "*Sam could also contact town or parish council level below the council*" [Participant 7]. While another highlighted that she should "*look at national data to see what else is available*." [Participant 5]. On the one hand these pushes towards alternative data sources may highlight an issue with the prompt, in that participants rejected the first step of Sam's story being contacting anyone. However, considering the importance of relationships in the

majority of the stories this could indicate underlying assumptions around the difficulty in accessing data within local government. Sam needed to have multiple approaches in order for her to maximise her chances of successfully obtaining it.

Theme 4: Ethical access to data

Discussion of the ethical requirements for data access was often a feature in the narrative. Participants mentioned "*data protection issues*" [Participant 8] or "*IG* [*information governance] permissions*" [Participant 1]. These aspects of the stories were related to formalised legal procedures within local government and not specific concerns related to moral justifications like privacy or consent. For example, some participants described steps within data sharing like anonymisation, privacy notices, data suppression, and secure transfer. Some highlighted that Sam may not be able to access the data due to these concerns, or may not be able to access the data at a finite level. These kinds of formalised ethics considerations were presented as both established working procedures and potentially as a roadblock toward data access. There was little to no mention of whether Sam *should* have the data or additional details described about the kind of data Sam would want. Ethics were characterised as things to be done and not necessarily as concerns about what Sam would do with the data. This highlights how ethics are often distilled into a set of procedures to overcome rather than more nuanced reflection on good practice and considering exactly what the ethical issues are as a precursor to determining what should be done (Haggerty, 2004).

Theme 5 and 5a: Data flows and the likelihood of an outcome

Many of the participants spent time writing about the practical components of a data project. This included establishing what kind of data was needed through conversations with relevant leaders in the council and whether data was available. As well as the specific fields of data, e.g. which variables in locally held datasets were needed and the need for a data sharing agreement with the council. For example, one participant stated that "*Sam will want to understand what data the council holds and the terms under which it can be shared and used*." [Participant 9]. Several participants also discussed whether Sam would need a third party to

mediate data access, for example an impartial local advocate who understands data access procedures. Many of these features can be summarised as describing the expected data flow from the council to Sam and how Sam would best navigate this flow of information. However, in describing these flows, several participants questioned the likelihood of a successful outcome. Participants used terms like 'hopefully' and 'probably' to describe whether Sam would eventually receive data. Some participants questioned whether data would even be useful to Sam, for example "*Sam will need to identify if the data can answer her questions*" [Participant 9]. While others questioned whether Sam would have the skills required to use the data, one participant highlighted that Sam would '*need help accessing/navigating websites*." [Participant 3]. These descriptions that data may not be available or that Sam would not have the skills to use it reflect underlying assumptions about the kind of people who ask for data from local government. Sam was positioned as an outsider, someone who did not have the status required to effectively navigate government websites or use data. It suggests that access to data is dependent on a contingency of factors including not only who Sam contacts but what her status as an individual is in relation to the council. This further adds weight to how these kinds of data provision projects are seen as complex and potentially problematic.

5.4 Discussion of results and story completion

5.4.1 How do the stories compare to other literature on data sharing in local government

Participant's stories highlighted a range of perceptions of data sharing in local government. Whilst we would not wish to draw broad conclusions based on a pilot study, we have identified themes related to the importance of effective working relationships, the motives behind data sharing, and the complexities of data flows. Overall data projects were typified by a sense of reticence to data sharing. The pathway to accessing data consisted of a range of often informal procedures that were difficult to design or specify ahead of time. The theme related to ethical access to data, and the potential for data protection guidelines to either enable or hold back data sharing is echoed in other literature. Malomo and Sena (2017, p. 12) contend that "[f]ragmentation of functions and competences

among different organizations in the public sector results in different interpretations of what can be shared and what cannot." Similarly, in a quantitative study of data sharing within US city governments, Welch, Feeney, and Park (2016, p. 399) found "[r]ule bound agencies are significantly more likely to share data....possibly because stricter rule orientation allows the organization to effectively clarify sharing needs and limitations." In other words, clear guidelines on what can and cannot be shared enables more frequent data sharing. The theme related to aligning organisational and personal priorities can also find support in other author's essays on government and data. Malomo and Sena (2017, p. 21) argue from a technical standpoint that "data projects have to be aligned with the strategic priorities of the organization so that support from the senior management and key stakeholders can be easily gained." The stories suggest a stronger political motivation behind aligning the priorities of the individual with the organisation, in that data sharing is contingent on expressing similar goals.

Some potential contingencies of data projects discussed in the literature that did not resonate in the stories included the need for adequate technological resources and data skills gaps as well as the need to make data sharing an institutional priority (Keller et al., 2012; Malomo & Sena, 2017; Milakovich, 2013). Keller et al. (2012, p. 7) describes this as having "a data-driven mindset". However, participant's discussion of Sam needing to align her priorities with that of the council demonstrates tacit recognition of the role of the 'institutional mindset' in data sharing and particularly how the council is positioned as a gatekeeper in local data projects. Outside of the literature on data usage, there is of course a detailed literature on the uptake of technologies for public administration including digital techniques like e-participation (Orange, Elliman, Lian Kor, & Tassabehji, 2007; Sivarajah et al., 2015). Themes found here like the need for a personal connection resonate with the key role of 'people' described in e-government adoption by Orange et al. (2007). Similarly, in literature around the history of local government, data sharing exists in a complex tradition of local public administration being critiqued for not modernizing (Orr & Vince, 2009).

The results found here offer reflection on some contingencies specific to data including challenges in data flows, data access, and ethical guidelines. Further research could explore how data technologies do or do not differ compared to adoption of previous technologies in local public administration.

5.4.2 The viability of story completion in uncovering data rhetoric

There were two different anticipated responses to the story completion exercise. The first was that participants' social discourse of data sharing would result in an affirmation of the positive rhetoric around data, e.g. they would show Sam successfully receiving data with little reflection on the contingencies of data sharing. The second scenario was that these stories would open up a range of both positive and negative aspects of data sharing in local government. Considering the range of candidate themes above, the exercise was successful in exploring perceived contingencies related to data sharing. Of particular note was the discussion around aligning priorities and questioning the likelihood of Sam receiving data. It is the first time, to our knowledge, that an empirical study has reflected the role of institutional priorities in shaping data projects in local government. As well themes highlighting the complexities of data access resulting in the need for alternative data sources are also novel. In addition to successfully kindling critical discussion of data sharing, the method also involved a strong element of creativity. This kind of creativity encourages participant engagement in responding to research questions. A qualitative approach is an essential alternative to studying government data use which is typically dominated by quantitative and post-positivist perspectives. Story completion has the potential to introduce nuance and therefore rigour to the way that data use is understood in local government. Story completion allows for research through hypothetical scenarios thus investigating both formal and informal practices in data usage, as well as allowing individuals to draw both on successes and failures in creating stories. It offers the potential to explore a vast range of processes that guide decision-making under the increasing prevalence of the civic datafication model.

Lastly, we consider the nature of the participants in this study and how this may impact their story responses. They were local managers and administrators as well as data analysts whose practices and responses would likely align with their experiences and interactions with data. The focus on aligning priorities and data ethics requirements, in particular, may relate to their experiences with conducting local data projects that likely involve routine practices like processing data ethics forms or freedom of information requests. Therefore, while these themes add to our understanding of data sharing practices of local government, they may and indeed likely do not represent the views of individuals in diverse and more politicised context, e.g. among politicians and policy-makers.

5.4.3 Limitations and recommendations for story completion

Story completion, as a method, is not suitable for all kinds of research questions. It poses hypothetical situations as a way to safely elicit beliefs and perceptions, practices and procedures. Responses will necessarily reflect the specifics of the story stem provided. In our case study, the story stem discussed Sam requesting data. While we attempted to make the story easy to continue regardless of technological expertise, this necessarily made the stem relatively simplistic. In particular, participants immediately discussing Sam circumventing contacting anyone and instead looking for alternative data sources may reflect rejection of the prompt. We also did not require any particular length of story nor time taken to write, in future we would ask participants to take at least five minutes to complete their stories and provide a longer prompt. We conducted the story completion pilot alongside semi-structured interviews, we therefore asked participants to complete a written form of the story. Several participants commented that they were not used to writing by hand and that may have resulted in a shorter story length. There would likely be longer responses if the story was completed verbally or online. In future, we would like to expand on this story prompt in an online study to compare results. The strength of our study was its use of a technique novel to both political science and social data research. As a first step to extend the use of story completion beyond psychology and business

research, we believe this has been useful in expanding the kinds of critical research done on how data technologies are shaping government processes.

5.5 Conclusion

Data use in local UK governments has been lauded as a way of improving government services and cutting costs. There is a strong social discourse both in government rhetoric and academic research that data use is the future of civic society. We discuss the potential for story completion to help unpick the contingencies of data use in local government. Story completion is the use of a hypothetical narrative stem to elicit discussion around potentially sensitive topics. Participants are encouraged to finish the story and the stories are then collated using narrative, discourse, or thematic techniques. In our pilot study of a story stem related to data sharing between local government and community workers we examined perceptions of data sharing. Key findings include the role of organisational priorities in determining data access as well as the opaque and complex pathways towards data sharing. Story completion presents a viable option for exploring both the facilitators and barriers of data projects in local government. Further research should examine how effective story completion is in a full study sample, as well as other government contexts. Finding methods of eliciting critical reflection on data use in government is essential to ensuring the rhetoric around building data systems that 'work for everyone' matches reality.

REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research : a practical guide for beginners*. London, UK: SAGE.

Bretherton, I., Ridgeway, D., & Cassidy, J. (1990). Assessing internal working models of the attachment relationship; An attachment story completion task for 3-year-olds. In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention*. Chicago, IL: The University of Chicago Press.

Clarke, V., Braun, V., & Wooles, K. (2015). Thou Shalt Not Covet Another Man? Exploring Constructions of Same-Sex and Different-Sex Infidelity Using Story Completion. *Journal of Community & Applied Social Psychology, 25*(2), 153-166. doi:10.1002/casp.2204

Clarke, V., Hayfield, N., Moller, N., & Tischner, I. (2017). *Once Upon a Time …: Story completion methods*. Cambridge, UK: Cambridge University Press.

Cornford, J., Wilson, R., Baines, S., & Richardson, R. (2013). Local governance in the new information ecology: the challenge of building interpretative communities. *Public Money & Management, 33*(3), 201-208. doi:10.1080/09540962.2013.785705

Data Power. (2017). *Data Power 2017 Program: An international conference on (big) data & power.* Paper presented at the Data Power 2017, Ottawa, ON.

Government Digital Service. (2018). Gov.UK.

Haggerty, K. D. (2004). Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology, 27*(4), 391-414.

Hancock, M. (2016). *Data science ethical framework launch: Matt Hancock speech*. London: Cabinet Office Retrieved from https://www.gov.uk/government/speeches/data-science-ethical-framework-launch-matt-hancock-speech.

Jarrar, Y. (2017). What is the role of government in the digital age? *World Economic Forum*. Retrieved from https://www.weforum.org/agenda/2017/02/role-of-government-digital-age-data/

Jones, M., Shanahan, E., & McBeth, M. (Eds.). (2014). *The science of stories: Applications of the narrative policy framework in public policy analysis* (1 ed.). New York, NY: Palgrave Macmillan US.

Keller, S. A., Koonin, S. E., & Shipp, S. (2012). Big data and city living – what can it do for us? *Significance, 9*(4), 4-7.

Kettl, D. F. (2016). Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles. *Governance, 29*(4), 573-579. doi:10.1111/gove.12211

Kitchin, R. (2013). The real-time city? Big data and smart urbanism. *GeoJournal, 79*(1), 1-14. doi:10.1007/s10708-013-9516-8

Kitzinger, C., & Powell, D. (1995). Engendering infidelity: Essentialist and social constructionist readings of a story completion task. *Feminism & Psychology, 5*(3), 345-372.

Malomo, F., & Sena, V. (2017). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. *Policy & Internet, 9*(1), 7-27.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753-1760.

Manzoni, J. (2017). *Speech: Big data in government: the challenges and opportunities*.

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business: Report on a research agenda setting workshop*. Retrieved from Washington, D.C.:

Mikic, N., & Terradas, M. M. (2017). Understanding Maternal Mentalizing Capacity and Attachment Representations of Children With Reactive Attachment Disorder: Two Case Illustrations. *Psychoanalytic Psychology*. doi:10.1037/pap0000153

Milakovich, M. E. (2013). *Anticipatory government: Integrating big data for smaller government*. Paper presented at the Internet, Politics, Policy 2012, Oxford.

Moore, S. M. (1995). Girls' understanding and social constructions of menarche. *Journal of Adolescence, 18*(1), 87-104.

Orange, G., Elliman, T., Lian Kor, A., & Tassabehji, R. (2007). Local government and social or innovation value. *Transforming Government: People, Process and Policy, 1*(3), 242-254. doi:10.1108/17506160710778086

Orr, K., & Vince, R. (2009). Traditions of Local Government. *Public Administration, 87*(3), 655-677. doi:10.1111/j.1467-9299.2009.01770.x

Sivarajah, U., Weerakkody, V., Waller, P., Lee, H., Irani, Z., Choi, Y., . . . Glikman, Y. (2015). The role of e-participation and open data in evidence-based policy decision making in local government. *Journal of Organizational Computing and Electronic Commerce, 26*(1-2), 64-79. doi:10.1080/10919392.2015.1125171

Ştefan, C. A., & Avram, J. (2017). Investigating Direct and Indirect Effects of Attachment on Internalizing and Externalizing Problems through Emotion Regulation in a Cross-Sectional Study. *Journal of Child and Family Studies, 26*(8), 2311-2323. doi:10.1007/s10826-017-0723-7

Taylor, A. S., Lindley, S., Regan, T., Sweeney, D., Vlachokyriakos, V., Grainger, L., & Lingel, J. (2015). *Data-in-place: Thinking through the relations between data and community*. Paper presented at the HCI for Civic Engagement, Seoul, SK.

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology,* [under review].

Ward, J. S., & Barker, A. (2013). *Undefined by data: A survey of big data definition*. University of St Andrews. St Andrews, UK.

Welch, E. W., Feeney, M. K., & Park, C. H. (2016). Determinants of data sharing in US city governments. *Government Information Quarterly, 33*(3), 393-403. doi:10.1016/j.giq.2016.07.002

LINKING TEXT

Chapter 5 found that local government data sharing was contingent on organisational aims, personal relationships, and unclear ethical guidelines. Despite the positive social and organisational rhetoric around the need for greater and more open data use, in completing their stories participants expressed hesitance toward assuming data sharing would take place. This is partially due to the nature of the participants included, e.g. local data analysts and managers, that likely have experience of data projects failing. Therefore, the social narrative they describe is one of challenge and political complications at the operational level rather than larger social narratives around policy-making aims related to data use. This chapter adds to the picture of civic datafication by demonstrating this political nature of data sharing at the operational level. Data collection and analysis are not just objective tools to inform decision-making but are located within complex political and ethical motivations. In order for public engagement to feasibly take place within the context of local government data usage, engagement exercises must reflect the complex pathways through which data is used in government. From a feminist perspective, this chapter allows for reflection on how data analysis is operationalised at the local level to encourage further public interaction. In particular, the challenges of aligning priorities and ethics guidelines demonstrate spaces for furthering public engagement. Public engagement could offer an opportunity for local government to better refine which and in what contexts data projects benefit local communities through clearer shared guidelines about what data could and should do in local communities.

Chapter 6, the sister paper to Chapter 5, uses the same participant sample to further explore the nature of government data use in relation to Objective 2. In this chapter semi-structured interviews are used to explore the barriers and facilitators of data projects in local UK government. Similarly, to its sister chapter, this paper seeks to understand the nuances of data projects that do and do not succeed in order to reflect on the potential for public engagement in the context of government data use. This paper is currently not submitted for publication.

CHAPTER 6

"So it's everything, data is everything": A qualitative study of the barriers and facilitators to data science in local UK government

CHAPTER DECLARATION

This declaration concerns the article entitled:									
"So it's everything, data is everything": A qualitative study of the barriers and facilitators to data science in local UK government									
Publication status (tick one)									
draft manuscript	✓	**Submitted**		**In review**		**Accepted**		**Published**	
Publication details (reference)	N/A								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate predominantly executed the formulation of ideas, methodology design, work, and presentation of data. Formulation of ideas: ESR formulate the ideas (90%) with help from JB and HD (10%). Design of methodology: ESR designed the methodology and interview schedule (85%) with supervision from JB and HD (15%). Experimental work: ESR conducted the interviews (100%). Presentation of data in journal format: ESR wrote up the results (85%) with help and supervision from JB and HD (15%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						**Date**	18/06/2018		

6.1 INTRODUCTION

The use and combination of pre-existing and new forms of data, i.e. data science, has been lauded as a step towards revolutionising local government. Data science exists along a spectrum from small data projects using single administrative datasets to large syntheses of linked 'big data'. More than engaging with data science, local governments are pushed towards adopting data science under a model of civic datafication. As the Local Government Association (2018a) (LGA) contends, "use of data plays an increasing role in designing, delivering and transforming public services to improve outcomes and drive efficiencies within current financial constraints." Civic datafication therefore refers to the ways in which data and data systems are increasingly inscribed not only in government but in the ways that citizens interact with government from local to international contexts. Under the civic datafication model, data science is not just a technology defined for its objective scientific aspects but also for its transformative social capacity. Markus and Topi (2015, p. 3) describe this more complex definition of data technologies, in their consideration of big data, "as a cluster or assemblage of data-related ideas, resources, and practices".

While data science, and what we term civic datafication, are often discussed for their potential to improve and revolutionise local government, challenges exist in translating these theoretical improvements into reality. Several authors have highlighted that the private sector has already engaged with data science, and that government and specifically local government have fallen behind (Malomo & Sena, 2017; Milakovich, 2013). The goal of this paper is to explore the practice of local government data science and to consider the reasons for the mismatch with these theoretical imaginations. Specifically, in a series of interviews with policy officials, our aim is to explore the barriers and facilitators of data projects in UK local government. The paper develops as follows. We first discuss the context for data science in local government. We then describe our methodology and study design. We follow this by presenting the results of our interview study including discussion on how these interviews fit within or advance the larger literature on

local public administration and data science. Finally we conclude by reflecting on the strengths and limitation of our approach and the contingencies of the civic datafication model.

6.1.1 Local government and data use

Local government in the United Kingdom is responsible for a range of regional services including transport, waste management, social care, libraries, business licensing and education (Gov.UK, 2018; Local Government Association, 2018b). According to the LGA (2018b), around one million individuals are employed by local government in the UK. In recent years, local government has been increasingly responsible for devolved services characterised as a push towards 'localism' (Cornford, Wilson, Baines, & Richardson, 2013; Malomo & Sena, 2017). Cornford et al. (2013, p. 202) describe this as 'the idea that the quality of information or knowledge available to decision-makers decays with distance, that local politicians 'know their patch' in ways that the distant 'man from the ministry' cannot." Malomo and Sena (2017) add that local government is also increasingly shifting towards a commissioning model, where their role is as funder and sign-poster and not provider of services. In recent years localism and commissioning models have increasingly been underpinned by arguments around increased data use and data technology. Data is seen as a key part of better understanding local contexts, i.e. localism, and to tracking service effectiveness under a commissioning model. In the LGA's (2018a) own terms the use of data will help "design services around user needs", "drive efficiencies and public service transformation" and "engage and empower citizens to build their communities" among other things.

Improvements from data science projects are imagined in organisational capabilities, citizen interactions, cost savings, services offered as well as predictive analytics (Civil Service World & SAS Institute Inc., 2015; Malomo & Sena, 2017). For example, councils are linking together disparate data sets on police and transport data to better respond to emergency incidents, and also linking together

social care and health data to improve their understanding of citizen needs (Symons, 2016). Alongside a range of smaller and larger data linking projects, this enthusiasm for local data analysis is primarily exemplified in the 'smart city' movement. As Keller, Koonin, and Shipp (2012) emphatically asks about smart cities, "what social challenges can be solved with big data? The facile answer is 'all of them'" (p. 6). The smart city is typified by better use of data to inform and govern, "imagine knowing practically every detail about a city. The state of its infrastructure, its inhabitants, its environment are all known to you, to high resolutions in time and in space" (Keller et al., 2012, p. 5). Smart cities are ones in which data, citizen engagement, and digital technologies form the central tenets of local government improvement (Keller et al., 2012). Thus projects ranging from smaller data science linkages to larger overall transformations in the ways in which local services are run are conceptualised as part of the data revolution. In his comments on Local Government's fragmented and unclear adoption of data technologies, the former Secretary of State for Housing, Communities and Local Government stated in late 2017, "embracing digital is no longer optional. It's not...something you can decide not to do...the internet came for newspapers. It came for retail. And now it's coming for local government. We can't ignore the wave. We have to ride it" (Javid MP, 2017). Data science is thus seen as being at the forefront of local government transformation.

What is missing in the rhetoric around the inevitability and inevitable good of civic datafication is a critical analysis of local government's readiness for this future as well as a reflection on the social dimensions of data science. Some authors have cautioned that smart cities are subject to such ethical and practical concerns. Kitchin (2014) highlights the risk of depending so fully on software and the potential corporatisation of city governance as key city technologies are run by the private sector. At the smaller end of data science projects, there are also more traditional concerns related to surveillance, privacy, and representivity (Joseph & Johnson, 2013; Schintler & Kulkarni, 2014; Stough & McBride, 2014). Other authors have similarly been critical of government readiness to adopt data science

(Cornford et al., 2013; Milakovich, 2013). There have been some concerns raised in the literature as to why local government has not achieved the data revolution promised. These include data silos where data is not shared, concerns around privacy, underinvestment in technology, and limitations on the kind of data available to local government (Malomo & Sena, 2017). Some authors suggest there are more intangible concerns like a lack of 'data-driven mindset' (Keller et al., 2012, p. 7). However, these projects and reflections have rarely been empirically driven and few have studied in practice how local governments use and imagine data science. Local governments are under pressure to adapt more quickly to the civic datafication model; however the benefits are challenging to realise while the facilitators and barriers to such adaptation go under-studied.

6.2 METHODS

6.2.1 Design

This study used semi-structured interviews to explore the barriers and facilitators to data science in local UK government. The study site was a local authority in the South West of England. Two case studies were identified by a local government contact. The first was an internal data project which was designed and run within the local authority using pre-existing administrative data to answer a question related to education services. The second project was a community-local authority collaboration that used local data from the private sector to answer a question related to transportation. These two case studies offered examples of the kinds of data projects common to local government in the UK including the use of 'big data' in the second case study and secondary analysis in both. They also demonstrated examples of projects that originate within local government versus those that originate through external organisations. These different projects were included to reflect the diversity and range of data science within local government, and not as comparative cases.

6.2.2 Sample

We used a snowballing method to recruit interview participants. Key contacts in both projects were asked to identify the individuals who worked on the project. Subsequent participants were also asked to identify individuals. Data collection stopped once all individuals who had worked on the projects were interviewed. Fifteen individuals were recommended for interviews, of which we contacted 13 as two were identified to not have worked on the project directly. Two individuals declined to participate, and one did not respond to our initial request. The final dataset consisted of 10 individuals who were either local government workers or community data volunteers. In total 11 interviews were conducted, one participant worked on both projects and was therefore interviewed twice. This resulted in eight interviews on the internal local government project and three on the community-local authority collaboration project. The participants were part of government and technical publics who specialise in the use, regulation, and analysis of data. They were civil servants including policy managers, data analysts, and administrators, as well as data analysts who work along but outside of local government. This sample did not include policy-makers or politicians and the responses to the interview questions should be understood within these parameters.

6.2.3 Procedure and analysis

Interviews were conducted between January 2017 and May 2017. The interview questions covered two broad areas: participants were asked to describe the project in detail from start to finish, and were asked their thoughts on the use of data in local government. A full list of prompts is available in Appendix A. Interviews were recorded, transcribed and then analysed using Braun and Clarke's (2006) specifications for thematic analysis. This procedure includes first familiarisation with the transcripts, then identification of potential themes or nodes, followed by re-reviewing and grouping into themes. Braun and Clarke (2006) specify that this process is dynamic with potentially multiple phases of

reviewing and re-coding. We broadly looked for initial nodes related to the conduct of data science and data projects in local government, with a particular focus on facilitators and barriers. After an initial familiarisation of the data, nodes were grouped into themes. ESR conducted the interviews, the initial read and thematic analysis, while HD and JB assisted with synthesising the themes. All analysis was conducted using the qualitative analysis software NVivo 10. Quotes from participants are linked to their participant number (P1, P2 etc.).

Table 6.1: Thematic analysis results of 11 qualitative semi-structured interviews

Primary Theme Group	**Related sub-themes**
The landscape of data projects in local government	Data as administration
	Data as a narrative
	Data as conferring objectivity
Facilitators and barriers of data projects in local government	The ebbs and flows of data projects challenge timescales
	Challenges in the scope and kind of data available
	Unclear guidelines encourage hesitance towards data sharing
	Challenges in resources: skills and software
	Relationships are key
	The value of 'playing' with data

6.3 RESULTS AND DISCUSSION

An initial coding of the 11 interviews resulted in 253 nodes. After re-coding and synthesis this was narrowed to two prominent themes and nine sub-themes. As presented in Table 6.1, the two primary grouping themes are the landscape of data projects in local government and facilitators and barriers of data projects. Data projects exist within complex and contested spaces around local commissioning and service delivery decisions. Landscape sub-themes relate to this environment and how data is used in local government. These include data projects fitting alongside administrative process, data as a tool for constructing

narratives, and the usefulness of data as an objective tool for commissioning. Facilitators and barriers include themes that relate to how quickly or successfully data projects are able to complete their stated aims. These six sub-themes include that data projects ebb and flow, the challenges in the kind of data available, unclear data sharing guidelines, limited resources, the key role of relationships and 'playing' with data as a motivator for projects.

6.4 THE LANDSCAPE OF DATA PROJECTS IN LOCAL GOVERNMENT

6.4.1 Data as administration

Data projects were often described as a core part of the administrative function of local government. This framing was underpinned by a focus on the need to determine what kind of data the council held, what the profile of current service need in the population was, and the requirement to forecast need and determine future provision. These processes were a means to the end for data analysis to inform commissioning activity, often in contexts where there were limited resources. Data was therefore positioned as crucial to the day-to-day processes of local government work.

> *"Where are we currently at? What is the funding situation and where do we need to be headed?" (P7)*

> *"What you're doing is you're looking closely at where you can maximise budgets better. In order to do that you have to be able to look at which areas are being met and what that costs in relation to the data. Those are difficult discussions to have, so it's much easier if you've got some data behind them." (P6)*

There was an explicit assumption that data analysis, at least in part, would lead to better decision-making within the commissioning and administrative process. In fact, data was framed as an essential tool in making policy decisions with local government. And as local government increasingly moves towards a commissioning model, e.g. funding services rather than running them, data analysis is seen as key to both tracking the uptake and function of these services and deciding which to commission.

> *"Data feeds into that decision. It's not the only thing, you couldn't ever make it purely on what the data says, I understand that. But it has to be one of the major factors" (P7).*

> *"If you don't have the right data, you can't commission effectively so it's everything, data is everything" (P3).*

Data as a mainly administrative function suggests that the purported transformative capabilities of civic datafication in terms of reimagining the relationship between local government and citizens is under-developed. The adoption of data in local government is less about revolution and more about incorporating data analysis into current working procedures. In an analysis of the history of local government in the UK, Orr and Vince (2009) suggest that local government has a long tradition of discussing modernization. They describe this as being 'critical of the present and contain[ing] resources which can be drawn upon to challenge or disrupt' (p. 671). The rhetoric around civic datafication as revolution and the reality of data as administration thus fits in a history of local government looking to the future as a way to critique current practice. That's not to say data science does not have the potential to create change but rather than local governments have complex histories and ways of working that are not easily revolutionised. For now, data science is primarily a practice within existing administrative process.

6.4.2 Data as a narrative

Although, as illustrated by P6 above, data are used within the commissioning process, the analysis of these data were presented as part of a policy narrative about need and service provision. Although data is positioned as key to the administration of local government, nearly all participants highlighted its role as informing or negating a story. The value of storytelling and interpretation were evident throughout the data project process. Data projects were described as testing already existing narratives but were also part of re-building narratives after analysis finished. Stories influenced what kind of data analysis was done, which subsequently influenced how the stories were told.

> *"It's not so much the data, it's the narrative that goes with it and getting that narrative out" (P1).*

> *"Data is used to tell stories, data is part of a cognitive process that people use to justify a position" (P2).*

Individuals who were better at fitting data into a narrative were seen as successful in persuading people of the validity of their narrative. Several participants discussed the importance of effectively communicating the final narrative in presentations. In fact data analysis was seen as useful regardless of whether they were used to support or refute policy narratives. Rather it was about using the data to construct and communicate a narrative.

> *"That really was a means to get the message out there and I think it did hit home." (P1).*

Civic datafication including data science as a narrative-making process fits within an established history of policy-making as a contested and complex process of narrative creation (Cairney & Heikkila, 2014; Stone, 1989). Commissioning decisions are ultimately political choices and authors like Stone (1989) argue that these choices are about creating and justifying a narrative about a particular policy problem. Thus data science is one more tool in this process. The role of data science in narrative creation in commissioning has been discussed by other authors, Kettl (2016) goes as far as suggesting that without a good narrative, good data analysis is useless. Thus we find here in local government that the perspective of policy-making as narrative creation plays out in how data analysis is used to evaluate and decide on commissioning options. Data analysis is a tool for establishing, testing, designing, and presenting narratives.

6.4.3 Data as conferring objectivity

Several participants described data as conferring objectivity on policy narratives; data provided a perceived unbiased and unemotional justification for commissioning choices. In fact, data was positioned as a tool for overriding other aspects of decision-making. Data analysis results were described as neutral safe

objects to have challenging discussions around. Data results were seen as unemotional and in many cases as self-evidently 'the truth'. It allowed commissioners to argue that they were separating out the ideological and emotional dimensions to policy choices. Put another way, 'data' was a neutral object that facilitated more expedient, less contested, commissioning decisions. Data could be used to help calm and limit discussions that were emotionally-driven. Participants highlighted that using data, while still only part of those conversations, was nevertheless the incontestable part.

> *"Everyone can kind of feel a bit safer with some breakdown of data. You know you can share the information more, to some extent it's not common language, but it is nonetheless, less emotional than some of the other work that I was getting, you know conversations I was having." (P6)*

> *"It should be black or white to a certain extent." (P7)*

While participants often discussed data as objective, emerging scholarship within critical data studies focuses on the subjective and political elements of data and data analysis. "Data may seem to be a nonpartisan term but, in fact, it is hardly neutral when it comes to impacts of decisions about what type of information should be collected, interpreted and released for use during policy-making processes" (Milakovich, 2013, p. 9). How, what, and why data are collected can be, and often is, socially and politically motivated. Participants using data as a tool to shut down opposition or calm emotions points towards a perception and social discourse that data is objective but a usage of data that is profoundly political. Participants imagined data to be neutral and objective but used it as a tool for political purposes. Data analysis may be a mostly administrative activity in that the analysis itself follows the current activities of local government, but the use of the analysis results and the design of data studies includes complex political processes of narrative creation and negotiation. Data science is a political tool.

6.5 FACILITATORS AND BARRIERS OF DATA PROJECTS IN LOCAL GOVERNMENT

6.5.1 The ebbs and flows of data projects challenge timescales

Participants highlighted that data work consists of natural ebbs and flows, which can create challenges in finishing projects on time or indeed defining finite boundaries of individual projects. Nearly all participants expressed frustration with the length of time taken in the various stages of data analysis and interpretation. In particular, the data 'cleaning' phase where administrative data was edited to be usable for statistical analysis took a disproportionate amount of time. This reflected challenges in secondary data use where the individuals doing the analysis were often at a disadvantage in understanding the characteristics of that data.

> *"There were some problems with it so I had to go back and forth to different people and query the data, query why things I was seeing that I wasn't expecting". (P1)*

A partial result of these ebbs and flows is that project timelines were difficult to define and thus participants struggled to point to finite boundaries between the administrative functions of local government and singular data projects. Indeed, data projects were often described as iterative without clear aims. When participants were asked if the aim of their project was met, they often stated they saw the individual data project as part of a larger system where the aim was yet to be met. This reflects how data projects are seen more as a tool alongside the administrative function of government rather than an independent piece of work. Data projects are therefore not easily confined to one particular question or set of work, but rather iterate and evolve alongside both the day-to-day functions of local government and the commissioning process.

> *"The trouble is, it's very difficult for me to distinguish [name of project 1], this particular project from everything else we've been trying to do with [name of policy area] and pushing it forward as something that we really need to get to grips with." (P5)*

> *"To conclude almost the whole thing, I mean you'll never finish it, it will never be finalised". (P5)*

Malomo and Sena (2017, p. 21) contend that "data project[s] must be well defined and their main objectives have to be realistic and shared among the

different parties involved in the project." While this would facilitate more successful data projects, the iterative nature of projects identified here suggests that it may not be realistic for local government to create finite boundaries around data projects. If anything, in addition to the more innovative forms of data technologies seen in the discussion around smart cities, it would be useful to think of how to encourage better integration of data projects into more traditional administrative procedures. For example, Piscopo, Siebes, and Hardman (2017) demonstrate the use of machine learning to reduce the number of variables that local government need to predict community capacity. These kinds of projects using new and research-based technologies that work alongside current practice may be more likely to succeed than projects that focus on separating out data projects from the procedures of local government.

6.5.2 Challenges in the scope and kind of data available

The quality and form of data available was seen as a major feature of successful data projects. What constitutes 'good' data was described by a range of factors including how well-organised the data was, whether it was representative, whether it was easily accessible and the cost of access. In terms of organisation and accessibility, data problems were often identified when data was aggregated therefore not individual or when data was messy and unorganised. Accessibility challenges included a lack of digitised administrative data as well as organisational challenges between different government departments that led to data access being siloed. For example, it was seen as particularly challenging to gain access to health data.

> *"The data's very old, some of the information we were looking at was all over the place, in terms of spreadsheets, the database is awful, so the work really needed doing in a more timely way." (P3)*

> *"So from a dataset, you can't work with it. It's like thank you very much but so what?" (P9)*

> *"We do tend to put in lots of different barriers, because you work for X and I work for Y and they work for Z." (P5)*

The concern around data quality for local government civic datafication is echoed in the literature (Civil Service World & SAS Institute Inc., 2015; Joseph & Johnson, 2013; Keller et al., 2012; Malomo & Sena, 2017). Clearly, it would be impossible to adapt to civic datafication without accessible and reliable data. Similarly, local government has a history of fragmentation in services which is reflected in the siloed nature of data collection and accessibility (Orr & Vince, 2009). While it is a familiar message that without good accessible data there can be no good data analysis, it is important to note the continued challenges that local government face in adapting to the call for innovation while also balancing complex, and at times outdated, data collection procedures. It will be impossible for local government to experience the recommended data science revolution without first experiencing a data collection revolution. Under the civic datafication model, such a revolution would need to fundamentally reassess what should be collected and the kinds of information that would best help improve the quality of services and the relationship between publics and local government.

6.5.3 Unclear guidelines encourage hesitance towards data sharing

While data sharing was discussed as a way of adding value to existing data by linking across services, unclear guidelines towards what constituted ethical data sharing created a hesitancy to challenge data silos. Unethical data sharing was seen as an organisational risk, particularly to data analysts working in partnership with local government. Formalised data sharing guidelines like information governance and data protection procedures shaped what kind of data projects would be done due to these sharing limitations. Information governance interpretation was thus a key process within data projects.

> *"There's various points that you'd go through, the first is suck through your teeth like you're a mechanic that's just seen your big end's gone and talk about, "Oh dear, that's going to be difficult" and that's largely from the Information Governance side of things." (P8)*

> *"Personally sensitive data and given that personal data issue, I wouldn't have wanted to take on the raw data, even if it was on offer because it adds a liability to us." (P10)*

These unclear guidelines contributed to the maintenance of data silos. This reflects the complexities described in working between sectors on a single data project. Participants struggled with negotiating how to access data in other departments in a way that did not breach the original purpose of the data collection as per information governance guidance. The accessibility of health data was often highlighted as a barrier to linking data.

> *"One of the things that we're still seeking to do to a certain extent is to break down the barriers to data sharing with the different stakeholders." (P5)*

> *"They want to be able to share the data with us, [they] just want to make sure that they're allowed to do it." (P4)*

Other authors have argued that unclear guidelines to data sharing and existing data silos make it challenging for government to incorporate data in decision-making. Malomo and Sena (2017, p. 12) argue that "[f]ragmentation of functions and competences among different organizations in the public sector results in different interpretations of what can be shared and what cannot." Similarly in a quantitative analysis, Welch, Feeney, and Park (2016, p. 399) found "[r]ule bound agencies are significantly more likely to share data with other government actors...possibly because stricter rule orientation allows the organization to effectively clarify sharing needs and limitations." Data silos present both a key barrier and an opportunity for local and national governments to improve the effectiveness of data in decision-making. Offering clearer, organisation-wide rules on what can and cannot be shared would be a major boon for data science to succeed within local government.

6.5.4 Challenges in resources: skills and software

Participants also discussed more practical concerns related to completing their projects including the limitations of having access to finite skills and software. Within the council, participants described a scarcity of data analysis

skills. This scarcity meant that the analysis was often done by individuals who were external to the team. Consequently, it created further challenges for analysts as they were least familiar with the data collection procedures. It also presented challenges for individual departments to adapt to data-driven commissioning procedures as they needed to outsource their work.

> *"Yeah, well I have one friend who's quite data crazy and she's not in my team. She's a colleague, but she's not in my team. I constantly realise the limitations of my team as a result of not having someone like her." (P6)*

> *"I've got those [data] skills which are probably fairly unique" (P1)*

Beyond the need for more data analysts, participants highlighted the limitations of software. Of note, participants external to the council had challenges in working between their own computers and the council's computing facilities. This created strong barriers to collaborative work, which is a key tenet of the smart city movement in the ability to link together services (Keller, 2012). Participants within the council also highlighted that the software they worked with was outdated and difficult to use.

> *"I tried to show him the [data] tool, and it was like oh, this is blank, and I could show him on my phone but I couldn't show it on his desktop. It was like, oh." (P9)*

> *"[An] old creaking [name] database...which hasn't been maintained properly for a while and yeah, getting the data out in any meaningful way is a nightmare, the software co who supply it are not easy to work with." (P3)*

While resources in terms of data analytic skills and effective modern software may seem like a fairly simple barrier to take down, the ability for local government to address this will, as always, be constrained by the funds available to them. Lack of resource is not a new phenomenon for local government in the UK (Joseph & Johnson, 2013; Keller et al., 2012; Kitchin, 2014; Malomo & Sena, 2017). As central government pushes a localism agenda (Cornford et al., 2013), this challenge is likely to become more problematic. In response, there will need to be greater emphasis on providing the training and resources needed to make effective local decisions using data technologies. It may be less about encouraging

local government to stop ignoring the digital revolution as suggested by central government (Javid MP, 2017), but about providing the basic building blocks for doing so, e.g. the skills and resources required for data analysis.

6.5.5 Relationships are key

All participants described the key role of personal relationships to the success or relative failure of data projects. This was characterised as needing to know 'the right people' who often had a familiarity with the dataset itself, the data collection process, or the service. They may also have experience in data analysis or know where data is held within the council. It was key to know who held what data, who had access to what data, and how it was collected in order to successfully navigate subsequent stages of the data projects.

> *"It's knowing who to go to and when you've found that person" (P1).*

While effective working relationships would be key to any form of project, regardless of data use, relationships in data projects were essential to understanding the data itself. For example, if you did not know who collected the data then you could not know the characteristics of the data and subsequently the most effective way to analyse it. A key aspect of a successful relationship was effective collaboration. Therefore successful data projects, e.g. ones that met their aims or resulted in a commissioning decision, relied on tacit awareness of skills and individuals within the council and how to best manage the relationships between team members. This knowledge was often held by individuals within the council, hence a data analyst being already embedded within the council would significantly decrease the time needed to start and complete a data project.

> *"One of the most important things in terms of my priority when I started [on the project] was to work really intensively on the relationships" (P6)*

Reflection on the human aspects of technology is a common theme in local government innovation (Orange, Elliman, Lian Kor, & Tassabehji, 2007). We add to this literature by suggesting that effective working collaborations are a constitutive part of translating data technologies into effective policy

administration and decision-making. Without individuals embedded within local governments themselves who have data science skills but also knowledge of local practice and data collection, adapting to data science within local government will continue to face strong challenges.

6.5.6 The value of 'playing' with data

While data projects were most often discussed as part of administration, participants also reflected on their personal interest in data use. Several participants used the phrase 'playing with the data' to describe this interest. This opportunity to access new data and to informally examine it motivated creative data analysis. Beyond motivation, play also helped participants gain data skills. Gaining data skills was often discussed as an important benefit to participating in data projects. Playing with the data, beyond personal interest, was also about starting analysis and making meaning from the data. Playing was an essential first stage in understanding the strengths and limitations of the datasets and therefore enabled analysts to query and examine any concerns with the quality of the data.

> *"Whenever we get any new data, there's lots of enthusiasm for playing with it" (P10)*

> *"Playing with data if you like, in looking at big spreadsheets and drawing any findings." (P9)*

While some of the barriers and facilitators mentioned above like challenging siloed working and adequate skills would be required for most kinds of cross-sectoral work within local government, others are unique to data projects. Particularly, the concept of playing with the data as a way to encourage innovative data exploration demonstrates the way that data science and civic datafication are unique. Embracing this playful aspect of civic datafication could, in fact, encourage more collaborative work both within government and between publics and government. If play is a positive aspect of data work, then opening local government datasets to enable play may encourage more and diverse publics to provide substantive input into data projects.

6.6 CONCLUSION

6.6.1 Final thoughts on civic datafication

The themes presented here run from highly political to highly practical. Some barriers are, evidently, easier to overcome than others. The nature of data as conferring objectivity and shutting down contested conversations demonstrate the deeply ingrained social and political nature of data science in local government. It is important to consider both in practice and theory how data science and civic datafication encourage use of data technologies to do these kinds of social and political things.

Local government is self-evidently located in close proximity to the citizens whom they seek to govern with data technologies. In a recent ODI and YouGov (2017) poll of 2,023 British adults, only 41% said they would trust local government to hold data about themselves. This apparent lack of trust is often assumed to be rooted in concerns around privacy and anonymization (Schintler & Kulkarni, 2014; Stough & McBride, 2014). However considering the nuanced and political nature of data discussed here, we would like to suggest a broader understanding of how citizens could interact with data in local government and how they may characterise their concerns. As Kitchin (2014, p. 9) critiques

> *"technocratic forms of governance are highly narrow in scope and reductionist and functionalist in approach, based on a limited set of particular kinds of data and failing to take account of the wider effects of culture, politics, policy, governance and capital that shape city life and how it unfolds."*

There are risks in fully adopting a model of civic datafication that does not account for the varied and political nature of data usage and civic life. Harms like those seen in US courts (Angwin, Larson, Mattu, & Kirchner, 2016) related to racially biased sentencing algorithms can and do occur. We echo the call made by Kitchin (2014) and others, as well as our own findings around the need for better consideration of data collections mechanisms, for data usage that is cognizant of public views and needs, particularly the ways in which citizens are engaged with the use of data (Milakovich, 2013). Keller et al. (2012, p. 7) imagines the "smart city

will be one where civic engagement is a reality: its people will participate in its running, its management, its ethos, and will feel part of that city." This, however, is not yet reality. The future of data projects in local government is closely and fundamentally connected to the social and political health of local communities. This closeness breeds a necessary inclusion of local citizens in how government use data.

6.6.2 Strengths, limitations, and critical reflection

This paper provides a rich, empirical qualitative exploration of the facilitators and barriers to local government data projects which is often relegated to purely quantitative or theoretical study. Our study raised relatively novel findings around creativity and play in data projects as well as how data is perceived to be an objective tool in narrative-making. Several of the themes found here could be explored for further empirical research as well as testing in practice. In particular, the themes related to data as conferring objectivity and neutral conversation would benefit from subsequent theoretical exploration. For practitioners, beyond accounting for the lack of resources and good quality data, providing space for innovative data exploration and play may encourage a greater uptake of data technologies in decision-making. Limitations include our smaller sample size and that participants were necessarily localised around two projects. However, considering the nature of our inquiry around lived-experience of data projects we feel that exploring these two projects in detail was an effective way to find nuanced understandings of common examples of local government data usage. As well, our sample included all participants in the two projects and expanding our search outside of those who worked on the projects would have been unnecessary. We also highlight that we may have missed key barriers to data projects as we only explored what could be considered completed data projects. These projects already faced the greatest barrier to any data project being completed, namely getting started.

The nature of the participants queried here e.g. local government workers who's experience with data projects will be at the operational level of collecting, using, and preparing data, likely influence the themes and discussion. For example, a lack of resources and the value of relationships are relevant to individuals who work at this operational level. Perspectives from national government and policy-makers/politicians could and would differ and may speak more strongly the social narratives around the need for local government to improve through further data usage as discussed in the introduction. However, the nature of the study was to understand these more practical and operational challenges of data analysis and thus the themes speak to that research aim. It is also important to note possible variations in the intensity and kind of data analysis in different local authorities across the UK. For example, resource challenges are likely a consistent problem across the UK but specific concerns around graphics cards and outdated software may be specific to this local authority. From a critical perspective, these themes should be understood within the characteristics of the study participants and the nature of the inquiry.

6.6.3 Conclusion

Local government in the UK is facing a data revolution alongside the development of devolved and localised services. Key to effectively navigating both of these phenomena is understanding and critiquing how local government adopts a model of civic datafication. We conducted a series of qualitative interviews with local government workers to explore the facilitators and barriers of data projects in local government. We found that the boundaries around data projects were often blurry as projects tended to ebb and flow. Key barriers included time, resources, and access to quality data. Participants also reflected on the role of data in creating and supporting narratives around commissioning decisions. Data was perceived as an objective tool to quiet contestations. Civic datafication has both great positive potential to improve local services and citizen engagement but also great risk in supporting technocratic and de-contextualised

decision-making. Critical exploration of how both local and national governments use data is essential to ensuring the former overcomes the latter.

REFERENCES

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. doi:10.1191/1478088706qp063oa

Cairney, P., & Heikkila, T. (2014). A comparison of theories of the policy process. In P. Sabatier & C. Weible (Eds.), Theories of the Policy Process (3rd ed.). Chicago: Westview Press.

Civil Service World, & SAS Institute Inc. (2015). Using big data to support a new government: Research survey report. Retrieved from https://www.sas.com/content/dam/SAS/en_gb/doc/whitepaper1/PublicSector/CSW-SAS-research-paper.pdf

Cornford, J., Wilson, R., Baines, S., & Richardson, R. (2013). Local governance in the new information ecology: the challenge of building interpretative communities. Public Money & Management, 33(3), 201-208. doi:10.1080/09540962.2013.785705

Gov.UK. (2018). Understand how your council works. Retrieved from https://www.gov.uk/understand-how-your-council-works

Javid MP, T. R. H. S. (2017). Changing technology, changing government, changing lives. Gov.UK Retrieved from https://www.gov.uk/government/speeches/changing-technology-changing-government-changing-lives.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. IT Professional, 15(6), 43-48.

Keller, S. A., Koonin, S. E., & Shipp, S. (2012). Big data and city living – what can it do for us? Significance, 9(4), 4-7.

Kettl, D. F. (2016). Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles. Governance, 29(4), 573-579. doi:10.1111/gove.12211

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, 1(1). doi:10.1177/2053951714528481

Local Government Association. (2018a). Better use of data. Retrieved from https://www.local.gov.uk/our-support/guidance-and-resources/data-and-transparency/better-use-data

Local Government Association. (2018b). What is local government? Retrieved from https://www.local.gov.uk/about/what-local-government

Malomo, F., & Sena, V. (2017). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. Policy & Internet, 9(1), 7-27.

Markus, M. L., & Topi, H. (2015). Big data, big decisions for science, society, and business: Report on a research agenda setting workshop. Retrieved from Washington, D.C.:

Milakovich, M. E. (2013). Anticipatory government: Integrating big data for smaller government. Paper presented at the Internet, Politics, Policy 2012, Oxford.

ODI, & YouGov. (2017). Attitudes towards data sharing - ODI/YouGov poll results - 2017-11-29 #PUBLIC. Retrieved from: https://docs.google.com/spreadsheets/d/1A_y1XioG2Y4gSy7wXE3kivE40ZiwXrpIbj-YujY_-CQ/edit#gid=471882920

Orange, G., Elliman, T., Lian Kor, A., & Tassabehji, R. (2007). Local government and social or innovation value. Transforming Government: People, Process and Policy, 1(3), 242-254. doi:10.1108/17506160710778086

Orr, K., & Vince, R. (2009). Traditions of Local Government. Public Administration, 87(3), 655-677. doi:10.1111/j.1467-9299.2009.01770.x

Piscopo, A., Siebes, R., & Hardman, L. (2017). Predicting Sense of Community and Participation by Applying Machine Learning to Open Government Data. Policy and Internet, 9(1), 55-75. doi:10.1002/poi3.145

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. Review of Policy Research, 31(4), 343-348.

Stone, D. A. (1989). Causal stories and the formation of policy agendas. Political Science Quarterly, 104(2), 281-300.

Stough, R., & McBride, D. (2014). Big data and U.S. public policy. Review of Policy Research, 31(4), 339-342.

Symons, T. (2016). Wise Council: Insights from the cutting edge of data-driven local government. Retrieved from: https://media.nesta.org.uk/documents/wise_council.pdf

Welch, E. W., Feeney, M. K., & Park, C. H. (2016). Determinants of data sharing in US city governments. Government Information Quarterly, 33(3), 393-403. doi:10.1016/j.giq.2016.07.002

LINKING TEXT

Chapter 6 found that government data projects are both political tools in that they are used to confer objectivity in difficult political discussions but also practically limited by the lack of skills and financial resources available to local government. Data projects were also not described in isolation to the other activities of local government, these projects had blurry boundaries between policy-making and administrative activities. The political nature of data use as found in the two previous chapters reflects the way that power infuses the relationships within civic datafication. For example, non-governmental publics were not discussed by the participants in the data projects, and while that may relate to the nature of the participants, it nevertheless speaks to the fact that publics are extremely limited in their involvement in data use in local government. Data analytics were a tool with which to shut down or open up policy conversations within local government. Therefore, publics outside local government face challenges in understanding and being involved in local government decision-making around data practices. This emphasises the need for better public engagement on the ways in which data is collected and used in local government with a range publics.

As mentioned in the chapter it is important to consider from a critical perspective how the discussions here are influenced by the nature of the inquiry and the participants. In different contexts, e.g. central government, with different participants, e.g. policy-makers, the challenges and political nature of data as described here may differ. However, local government provides an important target for public engagement from a feminist lens as it is non-governmental publics most proximal form of government-data interaction and should be seen as an important target for public engagement interventions. The preceding chapters also help better understand the nature of the power held by government in relation to data analysis. Power varies by the kind of government in question, and while it may be easy to critique the way that government conducts data analysis, it is important to note that local government does so under direction by diffuse and

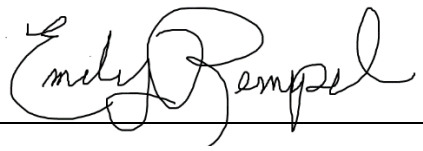
siloed government agencies at the local and national level, as well as under high resource limitations. Public engagement should reflect these challenges. This concludes this thesis' exploration of Objective 2.

Chapter 7 turns to address the final objective of this thesis, broadly to look forward to what form public engagement on government data use should take in the future. This answers the final dimensions of the relationships within civic datafication. Objective 1 focussed on the relationships between publics and government. Objective 2 focussed on the relationships between data and government. And finally Objective 3 focuses on the relationships between data and publics in the context of government. Chapter 7 uses thematic analysis of focus groups to compare and contrast differing public views on public engagement with local government data usage. The workshop materials in Chapter 7, including the focus group questions and the hypothetical public engagement online application, were designed based on the results of previous chapters. This included demonstrating that an online platform could host local data projects and better organise government data use which reflected the findings of limited capacity in Chapters 5 and 6. Using a feminist lens, the purpose of this chapter is to better include non-technical and non-governmental publics in deciding what good public engagement consists of. This chapter is not currently submitted for publication.

CHAPTER 7

Contrasting publics' views of public engagement on local government data use: A qualitative study

CHAPTER DECLARATION

This declaration concerns the article entitled:									
Contrasting publics' views of public engagement on local government data use: *A qualitative study*									
Publication status (tick one)									
draft manuscript	✓	**Submitted**		**In review**		**Accepted**		**Published**	
Publication details (reference)	n/a								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate predominantly executed the formulation of ideas, methodology design, work, and presentation of data. Formulation of ideas: ESR completed 80% of the formulation of ideas, with 10% help from an external RA in designing the survey, and supervision 10%. Design of methodology: ESR designed the workshop materials (85%) with help and supervision by JB and HD (15%). Experimental work: ESR ran the workshops and facilitated four out of seven focus groups (85%), she had help with facilitation from JB, HD, and two external researchers (15%). Presentation of data in journal format: ESR wrote the results primarily (85%) with editing and supervision by JB (15%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							**Date**	18/06/2018	

7.1 INTRODUCTION

Governments and citizens in the UK and around the world are entering a new age of data technology and data awareness. Terms like big data, data privacy, open data, and data protection are increasingly part of the public lexicon. Reports of data breaches and data harms like the revelations from Edward Snowden, the Cambridge Analytica influence on the 2016 US Election, and concern over the UK's Care.Data data sharing plan are now daily news (Adams, 2018; Carter, Laurie, & Dixon-Woods, 2015; Greenwald, MacAskill, & Poitras, 2013). In fact the European Union's new General Data Protection Regulations (GDPR) (European Parliament & Council of the European Union, 2016) proclaim that the "protection of natural persons in relation to the processing of personal data is a fundamental right". Data practices are increasingly thought of as socio-technical processes that must be debated by individuals, communities, and states.

The increased recognition of the importance of good data practice presents a challenge to private organisations and government to ensure that, as the GDPR (European Parliament & Council of the European Union, 2016) describes, the "processing of personal data should be designed to serve mankind." Governments in particular face a strong mandate to adhere to good data practice. However, this also challenges publics to reflect on who is and is not trustworthy to hold information about themselves both in government and outside of it and on a larger scale what sort of future they want. These kinds of decisions are dependent on contingencies of critical trust and diverse perspectives (Walls, Pidgeon, Weyman, & Horlick-Jones, 2004). At the community level, these are political and personal questions that as Davidson (2017, p. 3) writes in reflection on the policy process are "not only about means but also about ends. These are questions of what "ought to be"-questions of values, not facts." And due to the complexity of these kinds of questions around ethical and socially relevant data practice, and the ways in which circumstances around government data usage can alter trust, the answers are often not definitive. Subsequently, as the awareness and use of data technologies in government increases, so must the mechanisms through which

diverse and contrasting public opinions are enshrined in the governance of data. In other words, this paper advances the argument that in order to achieve a system of government data practice that serves 'mankind', public engagement must be an integral component to the conduct of that practice.

In order to establish a base on which to build a system of public engagement with data practice, our aim is to compare and contrast different public views on public engagement on this topic. The paper is structured as follows. First, we describe current literature on the practice of public engagement and data usage in local communities. Then we outline our methodology. We then present the results and a discussion of a series of workshops conducted with three different public groups and end with concluding remarks.

7.1.1 Public engagement and data use

Public engagement with data use, and more specifically with government data use, is a sub-discipline within the study of public engagement on science and technology. Public engagement is the inclusion of non-technical and non-governing publics in the development and governance of socio-technical, technological, and scientific objects (Braun & Konninger, 2017; Rowe & Frewer, 2005; Stilgoe, Lock, & Wilsdon, 2014). The activities under this umbrella range from highly critiqued communication exercises around technologies like nuclear power to collaborative exercises like citizen panels in the case of hydrogen energy (Flynn, Ricci, & Bellaby, 2012; Wynne, 2006). While there are less and more collaborative forms of public engagement, these activities as a whole are representative of the kinds of interactions between technical and non-technical publics. Public engagement is thus both a field that seeks to discover, understand, and establish the voice and viewpoints of publics in scientific governance but also to explore in what contexts, how effectively, and through which mechanisms publics are engaged. Public engagement study is both an activity of engaging publics but also of critiquing the practice of public engagement itself.

The topic of public engagement that this thesis examines is government data use, which covers a range of dimensions related to the design, collection, and

use of data in the activities of government. We do not focus on one particular data technology, for example machine learning, but rather the ecosystem around data usage. We term this ecosystem civic datafication, which is defined as the dramatic increase in the importance of data technologies to mediate the processes and relationships of communities, citizens, and government. These technologies include everything from tracking public health trends using social media data to algorithmic decision-making in criminal sentencing (Angwin, Larson, Mattu, & Kirchner, 2016; Poppy & Food Standards Agency, 2017). Data technologies can be applied both to the administrative functions of local government and to policy-making processes including designing, evaluating, and exploring policy options (Ginnis et al., 2016; Joseph & Johnson, 2013; Poppy & Food Standards Agency, 2017). These activities all aim to cut costs and drive efficiencies (Local Government Association, 2018). In this paper we focus on the use of these kinds of technologies in local government in the UK. Local governments face stark challenges in adapting to the technical and resource requirements of data technologies (Kitchin, 2013; Malomo & Sena, 2017). Subsequently, this offers an opportunity for public engagement to be of value for local government in helping to decide which technologies are best suited to their local area.

Government, to a minimal degree, have begun to consult publics on the ethical and social dimensions of data technologies (Cameron, Pope, Clemence, & Ipsos MORI Social Research Institute, 2014; Ginnis et al., 2016; The Royal Society, British Academy, Fourniol, & McLaughlin, 2017). A literature review, conducted by The Royal Society et al. (2017), of public engagement on data governance found that while publics were often limited in their knowledge of data technologies, they expressed a desire for data usage that was done in the 'public good'. The review ends by highlighting that more and better public engagement is needed in the context of data governance. Although these activities are a first step towards including the public in the debate around civic datafication, they are thus far one-off activities that do not involve ongoing deliberation and debate by publics on the topic of government data use. In addition, there has been minimal to no reflection

on the mechanisms of engagement in the context of data usage, nor critical evaluation of how to better integrate what the 'public good' and 'serving mankind' means in data practice at the local level or indeed any level (The Royal Society et al., 2017). This paper seeks to remedy this by engaging with three different kinds of publics to query what good data engagement looks like and indeed their preferences for such engagement.

7.2 METHODOLOGY

7.2.1 Design

This study aims to compare and contrast differing public views on public engagement with local government data usage through qualitative focus groups. The study site was a single local authority in the South West of England. The qualitative design included a task-based workshop with three different public groups.

7.2.2 Sample

Participants were enrolled from three different publics: local government officials, technical specialists, and non-technical publics. These are not indicative of all possible publics in relation to data but represent a broad spectrum of the kinds of individuals who may be engaged with government data practice. We primarily sought to identify pre-existing publics, e.g. those who already self-identified as belonging to one of the three groups. In the case of technical specialists we conducted three focus groups during a single workshop session with individuals who are part of a volunteer civic hacking group. For non-technical publics we enrolled groups who are civically active in the same local area as the technical group. This included one focus group/workshop with an environmental volunteer group as well as one focus group/workshop with a group of individuals recruited through a digital community message board. For local government, we conducted two focus groups within a single workshop session with those who work within the local authority that the other two publics reside within. In total we conducted seven focus groups over four workshops with 25 participants.

The local government group included mostly data analysts and policy area managers and administrators, therefore similar to the previous two studies these participants were not politicians. The technical group consisted mostly of data analysts or data enthusiasts while the non-technical and non-governmental groups were civically active but not data specialists. Their responses should be considered within these parameters.

7.2.3 Procedure

Each workshop consisted of a series of task-based discussions. Individual focus group discussions, which were held at different tables within the same room, were conducted during the workshops for the technical and government publics. The workshops lasted between one and two hours. All workshop materials, including focus group prompts, are provided in Appendix A. Each workshop started with an introduction to the topic area, e.g. government data use and public engagement, followed by a discussion of consent and anonymization in the context of signing the consent forms. This was followed by an activity that asked participants to identify where they created data in their day-to-day lives. During this task participants were asked to reflect on how government could use data and what constituted good and bad data usage. Participants were also presented with a mock-up of an online data engagement platform. This was used to encourage participants to talk about how they felt publics could be engaged around government data use. Finally participants were asked to fill in a workshop feedback form. Due to time constraints, participants in the non-technical public workshops were not able to fill in this feedback form.

7.2.4 Analysis

The workshop sessions were audio recorded and transcribed. The transcriptions were analysed using Braun and Clarke's (2006) method of thematic analysis. This includes iterative phases of familiarisation with the transcripts, initial coding of themes, categorisation of themes and finally review and re-categorisation. The transcripts of the three public groups were thematically analysed separately, i.e. the non-technical, technical, and government group

results are contrasted and compared. Two key questions were used to locate the thematic analysis. First, how do publics talk about data use in government? And second, how do publics discuss the inclusion of publics in government data use? Quotes are presented below where the participants are separated by focus group, where 'FG1' to 'FG7' refers to participants in one of the seven focus groups. 'I' refers to the interviewer/facilitator of that group.

7.3 RESULTS AND DISCUSSION

Presented in Table 7.1 are the results from the three workshop groups. Related themes are grouped together on each row by topic. This grouping is not based on agreement, i.e. in some cases themes are grouped together where the same topic is discussed but different viewpoints and perceptions are described. Following this table, the thematic analysis for the non-technical, technical, and government publics is presented in further detail. Next, the results from all groups are contrasted and discussed. And last, reflexive discussion on the workshops is presented.

Table 7.1: Thematic analysis results

	Workshop Group		
	Non-Technical Public	**Technical Public**	**Government Public**
Themes related to government data use	Participants discuss data in terms of its potential to do harm but also believe in the potential for it to benefit local communities.	Participants describe complex behaviours and a critical trust involved in deciding what was and was not a personal risk in relation to data.	Participants express personal fears related to data that results in a critical trust in organisations and safeguarding behaviours.
	Participants are critical of government data use based upon a perceived lack of change in their local area.	Participants question whether government uses data, while also being open and supportive of the use of data in government.	Government data use fits within a complex administrative culture where capacity is limited but data analysis is prized.

	Data technologies have changed the way the participants perceive the risks of their behaviours, particularly with commercial organisations.	It is impossible to separate out commercial involvement and the discussion of government data use in the context of data risks.	Data use in government exists within an inseparable ecosystem of both commercial and government data.

Table 7.1 ***cont'd***

Themes related to public engagement on government data use	Good data engagement empowers publics, recognises differences in the ability for individuals to engage, and offers multiple venues for engagement	Good data engagement is flexible, problem-focussed, and guided by intermediaries between publics and government.	Good data engagement is timely thus influential, is purpose-focussed, and of benefit for the organisation in question.
	Challenges for the success of data engagement include whether trust exists between publics and government and perceived public disinterest in civic activism.	Public engagement should be about empowerment, but empowerment may be a challenging and dangerous precedent to set.	Governments face risk in engaging publics are publics are imagined as combative and unknowledgeable about data use.

7.3.1 Non-technical workshop themes

Thematic analysis of the two non-technical workshops identified three themes related to government data use and two themes related to public engagement on government data use. First, participants often discussed data in terms of its potential to do harm, but this harm was underpinned by belief in the potential for data to benefit local communities and themselves. In many cases, data harms were hypothetical in that participants referred to examples of harms that were not rooted in their own experience. Many of these harms were also

related to the potential for data technologies to track and monitor individual behaviours.

> *FG2: But I think what just feels a little bit sort of, I can't think of the right word, I want to say sinister but that's not it, but …*
> *FG2, replying: I think that's the point, is it has been and people have been collecting your data…*

> *FG1: We would not have got this far if it hadn't been for that Facebook page. And part of me sometimes, like when it all came up about Cambridge Analytica, I just had to think I'm not deleting my Facebook page because I know politically this has served me well.*

Second, participants expressed a strong critique on government data use, this was not localised around harms or fears of government using data unethically but rather a larger critique about lack of change in their local area. There was a general sense of doubt expressed that government used data in any policy decisions. Participants struggled to point to examples of local government data use and discussed the need for greater communication on how local government uses data. Therefore, participants critiqued both the perceived lack of data use but also perceived ineffective local policies.

> *FG1: I guess my issue is that, and certainly what I found within transport, is that they have collected data, but there doesn't seem to be any thought about how they've collected some data, and then it doesn't actually apply to experience.*

> *FG2: Yeah, I don't think they are using it [referencing data].*
> *FG2, replying: No, no.*
> *FG2 replying: I really don't think they're using any information at all. And you look at the NHS and they're certainly not using it.*
> *FG2, replying: No, they're not, are they?*
> *FG2, replying: There's no joined up thinking there at all.*

Third, participants were quick to point to the ways in which data technologies have changed the way we perceive our behaviours and the risk associated with those behaviours. Participants highlighted that while day-to-day life may not have changed, in that we still do broadly the same things like

consume news, buy groceries, and talk to friends, the ways we do those things has changed immensely. For example, checking news on an app instead of a newspaper and ordering groceries online instead of going to a physical store. Many of the behaviours which resulted in data collection were related to the private sector, participant's views on data in general were closely connected to their perceptions of commercial activity. They highlighted that due to these changes, there is a need for public engagement that focusses on educating young children about data risks.

> *FG2: …every time I open up the app it suggests what I should be buying, so it's monitoring what I'm purchasing and where I'm purchasing it from, and it has my credit card.*

> *FG1: But this is a really good point. I mean education should be to help the children to navigate this, because how do they? It's such a completely different environment isn't it? I mean it's just so completely different. Even from when my children were growing up it's completely and utterly different.*

Fourth, participants described their perception of what good data engagement would look like. Specifically that engagement should empower publics, recognise differences in the ability for individuals to engage, and offer multiple venues for engagement. In terms of empowerment, participants expressed a desire to challenge decision-making in their local area as well as the kind of data that was collected. Participants also expressed that accessing public engagement activities should be easier, as they imagined other members of the public as uninterested in occupying the role of an 'activist'. Also they felt that others may find it difficult to engage with data technologies if they could only express their opinion online.

> *FG2: I'd like to feel is that there was a mechanism to allow my voice to be heard within that situation. So, OK, so I voted for person X who's now going to represent me, but person X needs to make the space or I need to make the space to actually explain what my views are, and that mechanism is often missing.*

> *FG2: They're the activists but I think there's a lot of people who sit sort of slightly on the fence who actually could be involved more if it was easier for them, that they didn't have to carry the banner and they didn't have to do that to feel that they wanted their voice to be heard. I do believe there's a place there.*

Fifth and lastly, participants discussed challenges for public engagement on local government data use, including whether trust exists between publics and government to the degree that trust may negate the desire for engagement in publics. The idea that trust in government means publics should not be engaged was not unproblematic. Participants disagreed on whether government officials could be relied on to make the best decisions for their local area. Another strong challenge was the perceived interest of publics. Participants, despite being members of publics they were describing, imagined publics to be disinterested in civic engagement. Thus participants, despite expressing strong views about the potential harms of data and the need for engagement, imagined the public as being unwilling to participate in public engagement exercises.

> *FG2: I think another point of view is that every four years we elect our councillors on their manifestos and having done that, we have authorised them to make these decisions on our behalf, as we do with our national politicians.*

> *FG1: You have to sort of actually want to do it don't you, rather than a lot of people just think, I just want to get on with my life, I don't really want to think about the council. I only want the council to fix the potholes.*

The themes around government data use exhibited a general sense of wariness and a critical trust of the datafied society among participants. This trust was contingent both on perceptions of data risks in the commercial sector as well as belief in whether government was addressing the issues of their local area. Cameron et al. (2014) have also found that government data use is considered more trustworthy by publics than commercial use. As well, trust is a common theme in public engagement literature (Stilgoe, Irwin, & Jones, 2006) and results here further add to the conceptualisation of trust as being critical and contingent on a range of events and interactions related to new technologies (Walls et al.,

2004). The discussion amongst the non-technical publics adds to this literature by suggesting that trust and engagement may be perceived to be mutually exclusive. Engagement was needed where there was not trust and vice versa. However, other participants challenged this and indeed described good engagement as substantive in nature, i.e. engagement that allowed publics to have influence over data-based decision-making in their local area (Fiorino, 1990). In sum, the non-technical workshops desired government to use data in a transparent, trustworthy way but to also engage publics in how they subsequently used that data to make policy decisions.

7.3.3 Technical workshop themes

Analysis of the technical publics workshop highlighted three themes related to the nature of government data use and two themes related to public engagement on local government data use. First, participants described complex behaviours and a critical trust involved in deciding what was and was not a personal risk in relation to data. Participants often discussed data re-identification and data tracking risks and described the ways in which they avoid these risks by controlling the data that is created about them. For example, by turning off GPS near their home address. There was a general sense of fatalism about data, not necessarily that data itself was harmful or would do harmful things, but that data would be collected and used regardless of personal choice. They also expressed a desire for greater transparency in how their data is collected and used in order to better evaluate what was and was not a risk. And that these risks could be tolerated, e.g. data sharing with third parties, if there was a clear line of personal or public benefit. For example there is a bargain to be made with companies like Google who provide digital services but as a result collect data through that service provision.

> *FG4: So it's one of those things that people really don't think about, what they see is a way of talking to friends what they don't realise they're also effectively selling away their, selling away their behaviour.*

FG5: So presumably that's transmitting where you are
FG5, replying: No, because I don't let it geocode
[laughter]
FG5, replying: It asks me where are you? And no because I'm in my house and I don't want it to know I'm in my house.

A second theme was the impossibility of separating out commercial interests and involvement in the discussion of government data use. Participants were quick to identify and discuss the ways in which the private sector use their data. This was related to the sense that the private sector had better adapted to the use of data technologies and subsequently would sell these technologies to government. Not only did they perceive their personal interactions with data collection and use to be tied to the commercial sector both in beneficial and harmful ways, they also perceived government's data collection and use to be tied to private companies.

FG3: So, I mean it's anything it's it's anything from loyalty cards in the supermarket where you've got to give your name and address isn't it?
FG3, replying: Yeah.
FG3, replying: It's websites...
FG3, replying: You agree then that they can use your information.

FG5: So I would like to think that government, local government would get more access to the data and do something with it but I kind of think it will be more, that data will be used to create…
FG5, replying: products
FG5, replying: …products to sell to local government…

In further discussing government data use, participants identified many ways that they thought local government use data, however they did not perceive that use to be highly effective. Participants often questioned whether government used data in any way and were highly critical of governments both local and national not employing scientific results in decision-making. However, participants were also quite open and supportive of the use of data in government such that one participant even described government as having a right to the data

of citizens. Participants were hopeful that as data technology advanced so would government use of data technologies.

> *FG4: I think any data well not any data from the government is okay, I think that they should have the right by from a principle to then to have access to the data of the citizens. That's my opinion, my personal opinion.*

> *I: How do you think local government uses data at the moment?*
> *FG3: That's interesting no [unintelligible] just ignore, just ignore the data*
> *I: They're ignoring the data? Okay that's interesting.*
> *FG3: [unintelligible] I don't, I don't think that local government or government is in any way scientific or data-based. Do you?*

Turning to the two themes about public engagement on government data use, participants perceived good engagement as flexible, problem-focussed and guided. By flexible, they described personal experiences with engaging publics in which they had to provide multiple venues of potential public participation. They highlighted that solely technical offerings like website-based data collaboration would be ineffective at encouraging participation. They suggested public engagement should recognise the availability, skills, and desires of publics to be involved. By problem-focussed they expressed that, in the end, engagement should be about the policy area in question and not necessarily about designing or using data. And finally that good engagement would be guided, in that publics would be offered specified and directed ways of engaging. They perceived themselves as intermediaries between government and publics and that their job was to do the more technological tasks in a data project.

> *FG5: The other thing is that building a website or an app to collect data isn't, isn't the hard part. The hard part is getting people to fill it in and actually go out and do the work-*

> *FG5: So it's kind of like identify the problem in the community and then that feels like a better hook-*
> *FG5, replying: And probably concrete things that they can do rather than I want to help uh…*

Finally, participants briefly discussed the value and purpose of public engagement. Specifically, they debated whether engagement should be about empowerment or if empowerment is a challenging and dangerous precedent to set. They expressed several concerns related to including publics who may not be data literate in data decision-making, including whether they would be able to accurately and correctly collect and use data. However, other participants also expressed that engagement should in fact enable publics to challenge decision-making and hold government to account in their data practice. These views are not diametrically opposed rather the three groups focussed on different aspects of engagement. Participants who were concerned about including publics held stronger views about data projects as technical exercises while participants who viewed engagement as empowerment focussed more on data projects as exercises in problem definition and policy-making.

> *FG3: So in a way you do want people to, but you can't, uh, what am I trying to say? You can't just have people collecting bits and bobs of data*
> *I: Right, okay, okay...*
> *FG3, replying: In a very unscientific way…*

> *FG5: I think one thing that I see as doing, it's doing, is we-we're empowering people to kind of act with data. So we're kind of acting like that human-interface for them.*

Overall participants in the technical workshop exhibited similar views of data use both within and outside of government compared to the non-technical publics. They also described holding a form of critical trust in data usage (Walls et al., 2004) as well as views of data that were strongly tied to their perceptions of commercial enterprise. While other public engagements have found this interconnection between commercial and government data use (Cameron et al., 2014), in their review of public engagements on data use, The Royal Society et al. (2017) concluded that the public held a "relatively unengaged attitude to data use by commercial organisations" (p. 3) as compared to the public sector. We find our results diverge with these findings. Instead, participants described a critical trust

in data which is contingent both on commercial and government data use. We did not find that participants were 'unengaged' on one or the other but rather participants drew on examples of both in discussing their fears and hopes for data use in government. This participant group was also highly critical of government and highlighted the need for data use that had a clear line of public or personal benefit. Needing a public benefit is common theme in public engagements around data use and is often described as doing data analysis in the 'public good' (Ginnis et al., 2016; The Royal Society et al., 2017).

Participants also expressed critical views of public engagement. They expressed concerns of data literacy and the capacity for publics to engage. Michael and Lupton (2016) similarly describe the challenge facing non-technical publics in engaging with data technologies. Capacity-building has also been referenced within public engagement scholarship as a key aspect of public engagement exercises (Selin et al., 2017). Proponents of a capacity-building model argue that the benefit of public engagement is learning about the technologies at hand. Gray, Gerlitz, and Bounegru (2018) also highlight the value that data literacy initiatives, particularly in their case on knowledge surrounding data infrastructure, can play in encouraging publics to engage with data. While data literacy is evidently important in empowering publics, participants here also highlighted that good engagement allows for a multitude of participation mechanisms. They described good engagement as allowing multiple venues and methods for engagement with data projects which has also been highlighted as a finding in public engagement around citizen science (Bonney, Phillips, Ballard, & Enck, 2016). In sum, the technical publics held similar critical views of government and data technologies but highlighted differing views on public engagement than the non-technical group. While both groups discussed public engagement in its ability to empower publics, the technical group added a general sense of doubt over the capacity of non-technical publics to engage in data processes.

7.3.4 Government workshop themes

Our analysis of the government workshop identified three themes related to views on government data use and two themes on public engagement on government data use. First that government data use fits within a complex administrative culture where capacity is limited but data analysis is prized. Participants highlighted that they perceived government to be limited in their ability to use data due to safeguarding requirements and the limitation of purpose, i.e. that data collected for one purpose could not be used for another. They perceived risks to individuals, in the context of safeguarding if data was identifiable, thus they were hesitant to support a large system of open data sharing. They also described administrative systems that were complex and outdated, and consequently in their experience it was challenging to gain access to data within local government. Despite these challenges data analysis was perceived as an opportunity provide benefit for citizens and make better decisions within government.

> *FG7: …so in the end I can't share that data, because there's no consent there, so I get a bit restricted. But it's probably not benefitting people, because it's not supplying the evidence they need to show they are working hard with those, but I'm restricted by the protocol. I find that that's quite hampering...*
>
> *FG6: But what was interesting in addition in terms of the use of that data is the assumption you state that approach, you think that's a big data data science job. It was a form filling job, because every single agency, every single one involved, although they were cheerfully putting all this data into 32 different IT systems, not one of them could get it out in a useful enough format for us to link it all together.*

Participants also discussed their personal perceptions of data, particularly their fears of data. These fears related to their trust in organisations and subsequent safeguarding behaviours at a personal level. Several participants were protective of their privacy and actively sought to not use social media in an effort to protect themselves. Certain organisations were more trustworthy than others, for example a dental or medical practice, and this was related to the institution itself and not to the data practices of that organisation. In that sense data sharing

was less risky with organisations that had a pre-established sense of reliability and trust.

> *I: What does in confidence mean to you? Let's start with in confidence.*
> *FG6: Secure and safe from...*
> *FG6, replying: Inappropriate.*
> *FG6, replying: There's always that sort of fear that it'll [data] be used.*

> *FG7: Maybe I don't know that their IT systems [dental practice] are any more secure than the IT systems here, I suppose I've just been brought up believing you have to trust at some point don't you? I suppose I would trust that...*
> *FG7, replying: Trust in the institution itself.*
> *FG7, replying: Yeah.*

> *FG7: Absolutely, I don't like social media or anything like that, because of it I don't do Facebook, I'm not on Twitter or anything like that. So, when I see things to say that what you're saying is going to be used and published and things, for me that just shouts alarm bells, because I don't want anything that I say to be out there. I'm quite private and keep myself to myself.*
> *FG7, replying: I agree, I don't use Facebook or Twitter or anything for exactly the same reasons, or post photos or anything. I don't want anybody ... I'm not doing anything exciting, but I don't want everybody to know.*

A third theme related to government data use, was that similar to the other two workshops, data use in government exists within an inseparable ecosystem of both commercial and government data. Views on one or the other are intrinsically linked. Commercial data collection was often discussed in relation to mobile phone usage and the way that being connected online related to fears of tracking and monitoring. Although this discussion was closely related to personal fears and concerns, participants expressed that the reason commercial tracking was more concerning is that it was not considered to be done in the public good. Thus data usage by the private sector was perceived to be about commercial benefit and subsequently less trustworthy.

> *FG7: I don't want people to use the information to try and sell things to me. If they're going to use it for commercial gain, then I don't want that.*

> *FG6: Google will tell you if there's traffic on the routes.*

FG6, replying: Telling my time of train, whether it's delayed.
I: How does it do this?
FG6, replying: I don't know. It's frightening.
FG6, replying: If you check the train times on National Rail...
FG6, replying: Not even that. They work out where you're stood. It's predominantly through location analytics.

In relation to public engagement, participants described good public engagement as being timely thus influential, purpose-focussed, and having a benefit for the organisation in question. Being timely related to ensuring public engagement happened before rather than after a policy decision was made which again related to whether there was an opportunity for publics to influence government decision-making. Evidently, in order to impact policy decisions publics must be included before a decision is made. Public engagement was discussed in relation to policy-making and not on data processes themselves, participants viewed public engagement as focussing on a political purpose. Public engagement and data sharing practices were also both described as needing to have a benefit to the individual departments within local government. Subsequently, organisations would not be willing to share data for public engagement or policy-making purposes unless they could see how it would be of benefit to themselves.

FG6: It's better the consultation doing it after the event where we don't take anybody's notice. [Laughs]

FG7: So, if there's a purpose, and there's actual things that you can measure, then it can be worthwhile. So, it's having that purpose to begin with I think.

FG6: This was a quote from one of our managers, what's in it for me? This was around releasing some open data. What's in it for me? How would I accredit that the person looking at the data knew what they were talking about?

Lastly, participants debated the risks for government in engaging publics. They imagined the non-governmental public as combative and, in some cases, unknowledgeable. Publics were not imagined as qualified to be involved in the

technical and scientific aspects of data collaboration. As such participants felt publics may misinterpret findings and that would lead to a worsening of the perceived combative relationship between local government and publics. Public engagement was discussed in many cases as an organisational risk. However, this idea was debated amongst the groups. Not all participants agreed that publics were unknowledgeable, or indeed that engagement would be combative. While the groups did not find agreement, these kinds of debates demonstrate that public engagement within local government data use must overcome both technical challenges related to data use but also the perceived negative relationship between the citizen and the state.

> *FG7: But if you're putting in lots of information out there, and people are going to be jumping to the wrong conclusions on different things…*
>
> *Fg6: I don't think we've got a very well educated society in this country, compared to some.*
> *FG6, replying: But we're part of the society. We're assuming the public are stupid, but we're the...*
> *FG6, replying: They are!*
> *FG6, replying: We're the public though. The public health jobs, they have skills, they have experience. I don't know, this is the conflicts of ... Well everyone can be stupid but assuming that they have no experience, expertise, knowledge, understanding, that's not true.*

Participants in the government workshop again described very similar views to the other two groups. They described a sense of wariness toward data and a view of government data use as insufficient. However, in contrast to the other groups, the government group highlighted why they felt there was a gap between the aspirations of government data use and the reality. They described a complex administrative culture that is not easily transformed by data technologies. This finding echoes true in academic discussion around government data use as lagging behind the private sector (Kitchin, 2013; Malomo & Sena, 2017). As Poel, Meyer, and Schroeder (2018, p. 17) state "policymaking and the data that underpins it lags behind the reality of the complex policymaking world

that it seeks to guide and transform." This group, like the technical public group, again highlighted the challenge of data literacy. Some participants perceived the public as being generally unknowledgeable. This likely stemmed from their perception and discussion of the 'general' public as combative. Government imagining the public as simultaneously difficult, i.e. combative, and disinterested ,i.e. unknowledgeable, has been found in other studies related to the views of policy-makers on publics (Barnett, Burningham, Walker, & Cass, 2010; Burningham, Barnett, & Walker, 2014; Laurent, 2007). The findings here resonate and add to this area of research by demonstrating that these imaginations of the public have a direct effect on how policy-makers subsequently wish to engage around data technologies. Public engagement with government data use is impossible within a perceived hostile relationship between publics and the state (Warburton, 2009).

7.3.5 Contrasting and comparing the three publics

Next we discuss broader themes related to what was and was not said by the three groups. A meta-commentary exists within these findings around how publics imagine themselves that builds on previous literature on the public imaginations of policy-makers (Burningham et al., 2014; Lezaun & Soneryd, 2007). As was pointed out by one of the government participants in this study, government workers are also members of publics. However, participants in all groups struggled to combine their imaginations of publics with their own beliefs and behaviours. In both the technical and government groups they imagined publics as unintelligent, and all three groups imagined publics as uninterested, despite the fact that they perceive themselves to be interested and intelligent. In the literature, policy-makers often perceive 'activist' publics as difficult and combative and non-activist publics as disinterested and unengaged (Burningham et al., 2014; Shelley-Egan & Davies, 2013; Torgersen & Schmidt, 2013; Walker, Cass, Burningham, & Barnett, 2010). This study adds that these stereotypes also pervade the imagination of publics by publics themselves. Publics imagine themselves to be at best disinterested and at worst too difficult to engage. There is

a concerning, self-fulfilling cycle in which the 'public' is perceived to be uninterested and therefore is not engaged.

This leads to the final question of this paper. What do publics both within and outside of government want in relation to public engagement on government data use? While publics were not necessarily imagined by the three groups to be interested in public engagement, they nevertheless discussed that public engagement should be empowering for publics. Publics should be engaged early in the policy process and allowed to influence the decisions that are made. Public engagement should also be flexible both in topic and form. Publics should be engaged both in how data is collected but also on the policy problem that government is trying to address with data. Public engagement was also valued in its potential ability to build capacity and knowledge of data processes in non-technical publics. These conceptions of good public engagement find support in the literature. Michael and Lupton (2016) and Gray et al. (2018) both discuss the need for capacity building amongst publics specifically in the context of data technologies. While Bonney et al. (2016) in their discussion of citizen science highlight the need for multiple venues and kinds of public engagement. Burri and Bellucci (2008) similarly argue for publics to be involved early in decisions made around new technologies. Unique to data technologies, however, is the need to focus both on the policy topic at hand and the substance of data technologies. While data technologies are socio-technical objects in and of themselves to be debated, in the context of government data use they also aim to do political things. Thus public engagement on government data use must focus both on means and ends. Overcoming the perception of publics as disinterested and unintelligent is evidently a challenge for public engagement on government data use. However, this is not necessarily a challenge for publics to evidence their interest and intelligence. Rather this is a challenge for government and technical publics to perceive of public engagement on broader terms. The use of data technologies is not only a technical exercise. As discussed here good public engagement on government data use must be flexible, influential, and offer publics multiple ways

of participation. This is best described by one of the participants in the technical workshops.

> *Just cause I think that, if you-if you want to get people involved then you have to be where the people are.*

7.3.6 Reflexivity, strengths, and limitations

The main strength of the study lies in its novel exploration of different public views on the same topic. This provided an opportunity to both explore differing perspectives while also building on feedback from these publics to create recommendations for public engagement on government data use. Rather than discussing limitations in a purist scientific sense, we reflect on the process as a whole to describe the ways in which the organisation of the workshops may have influenced the results presented above.

Some of the tasks were, on reflection, better suited to drawing out conversation about public engagement. When designing the workshops, activities were chosen that were relatively simple in order to be accessible to a range of individual data literacies. Although these activities were successful at easing into discussion around data technologies and public engagement some led to participants discussing relatively irrelevant topics. For example, the first exercise around discussing consent forms often led participants to express their thoughts about the university system.

The first workshop was recruited in conjunction with the civic hacking group, it led to some confusion for the participants on whether they were attending a normal meeting of the group or the research workshop. One of the participants had not opened the consent forms when they were emailed to them. They then came to the workshop expressing that they felt manipulated to attend as they did not know it would be a research session. Subsequently they did not participate in the workshop and the individual was not counted in the total number of participants. As well during the same workshop, two of the focus group recordings cut off the last 20-30 minutes. Participants in this group also questioned

why the facilitators needed to record their gender identity in the transcripts and on reflection the authors decided not to include a reference to whether the participant was male, female, or another gender identity.

In the government workshop, one participant attended who had also attended the civic hacking group. They had thought the session would be different, but indeed it was not. It was decided to let them participate as the nature of the group around the table was different. Thus they may have expressed different views. There was also a large challenge in recruitment for the final two focus groups with non-technical publics. Despite several stages of recruitment and a multitude of advertising, it was difficult to find individuals willing to participate. This may have been due to a lack of interest but also difficulty faced by the researchers in finding the avenues that individuals use to find events in their community. As a solution, the remuneration was increased which led to finding participants for the final workshop. The sum of these challenges meant that the workshops were iterative and flexible, and that discussion may have varied due to these challenges.

Lastly, we reflect on identity and presentation in the workshops. It was challenging to encourage debate on data usage without presenting examples of such data use. While there was an effort to highlight both good and bad examples of data usage, inevitably some of the examples and questions may have led to negative responses. Participant response also varied according to the identity of the facilitators and publics. The lead facilitator was a young, white, female who has a Canadian accent. This visible and oral identity may have influenced the ways that participants responded. In particular in the technical group, respondents were quick to offer advice as experts and to challenge whether the facilitator was qualified to discuss data technologies. In listening back to the recordings, the lead facilitator was surprised the number of times these participants gently mocked or challenged her qualifications to lead the sessions. However in the non-technical group, participants often referred to the facilitator as being an expert and deferring to her where they thought they did not know much about data technologies. They evidenced a hesitance to critique the

materials presented. These different responses within the groups and also towards the facilitators may have led the discussion either towards less critique or indeed more.

7.4 Conclusion

In conclusion, this workshop aimed to compare and contrast different public views on public engagement on government data use. We did this through seven of focus groups with technical, non-technical and government publics in a local authority in the UK. All three groups expressed a wariness towards data collection around their day-to-day activities as well as a critique of how well local government uses data in policy decision-making. The groups diverged in their discussion around public engagement. The technical and government groups focussed on their concerns around the 'public' as being data illiterate while all three groups expressed concerns around the 'public' as being disinterested in data use and policy-making. However, all three groups described good public engagement as focussing both on the means and ends of policy-making with data, being influential on local government process, and offering multiple mechanisms of participation.

REFERENCES

Adams, T. (2018). Facebook's week of shame: the Cambridge Analytica fallout. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Barnett, J., Burningham, K., Walker, G., & Cass, N. (2010). Imagined publics and engagement around renewable energy technologies in the UK. *Public Understanding of Science, 21*(1), 36-50. doi:10.1177/0963662510365663

Barnett, J., Cooper, H., & Senior, V. (2007). Belief in public efficacy, trust, and attitudes toward modern genetic science. *Risk Analysis, 27*(4), 921-933. doi:10.1111/j.1539-6924.2007.00932.x

Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Underst Sci, 25*(1), 2-16. doi:10.1177/0963662515607406

Braun, K., & Konninger, S. (2017). From experiments to ecosystems? Reviewing public participation, scientific governance and the systemic turn. *Public Underst Sci*, 963662517717375. doi:10.1177/0963662517717375

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Burningham, K., Barnett, J., & Walker, G. (2014). An Array of Deficits: Unpacking NIMBY Discourses in Wind Energy Developers' Conceptualizations of Their Local Opponents. *Society & Natural Resources, 28*(3), 246-260. doi:10.1080/08941920.2014.933923

Burri, R. V., & Bellucci, S. (2008). Public perception of nanotechnology. *Journal of Nanoparticle Research, 10*(3), 387-391. doi:10.1007/s11051-007-9286-7

Cameron, D., Pope, S., Clemence, M., & Ipsos MORI Social Research Institute. (2014). *Dialogue on Data - Exploring the public's views on using linked administrative data for research purposes*. Retrieved from London, UK: https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics, 41*(5), 404-409. doi:10.1136/medethics-2014-102374

Davidson, B. (2017). Storytelling and evidence-based policy: lessons from the grey literature. *Palgrave Communications, 3*, 17093. doi:10.1057/palcomms.2017.93

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), (2016).

Fiorino, D. J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. *Science Technology Human Values, 15*(2), 226-243.

Flynn, R., Ricci, M., & Bellaby, P. (2012). Ambiguity, complexity and uncertainty surrounding the hazards of hydrogen and public views of emergent risks. *Journal of Risk Research, 15*(4), 373-387. doi:10.1080/13669877.2011.634517

Ginnis, S., Evans, H., Boal, N., Davies, E., Palmqvist Aslaksen, A., Ipsos MORI Social Research Institute, . . . Sciencewise. (2016). *Public dialogue on the ethics of data science in government*. Retrieved from London, UK: https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society, 5*(2), 205395171878631. doi:10.1177/2053951718786316

Greenwald, G., MacAskill, E., & Poitras, K. (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelation. *The Guardian*.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional, 15*(6), 43-48.

Kitchin, R. (2013). The real-time city? Big data and smart urbanism. *GeoJournal, 79*(1), 1-14. doi:10.1007/s10708-013-9516-8

Laurent, B. (2007). Diverging convergences - Competing meanings of nanotechnology and converging technologies in a local context. *Innovation-the European Journal of Social Science Research, 20*(4), 343-357. doi:10.1080/13511610701760804

Lezaun, J., & Soneryd, L. (2007). Consulting citizens: technologies of elicitation and the mobility of publics. *Public Understanding of Science, 16*(3), 279-297. doi:10.1177/0963662507079371

Local Government Association. (2018). Better use of data. Retrieved from https://www.local.gov.uk/our-support/guidance-and-resources/data-and-transparency/better-use-data

Malomo, F., & Sena, V. (2017). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. *Policy & Internet, 9*(1), 7-27.

Michael, M., & Lupton, D. (2016). Toward a manifesto for the 'public understanding of big data'. *Public Underst Sci, 25*(1), 104-116. doi:10.1177/0963662515609005

Park, A., Curtice, J., Thomson, K., Bromley, C., & Phillips, M. (2005). *British Social Attitudes: The 21st report*. Retrieved from London, UK [SAGE Publications Ltd.].

Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big Data for Policymaking: Great Expectations, but with Limited Progress? *Policy & Internet*. doi:10.1002/poi3.176

Poppy, G., & Food Standards Agency. (2017). *Chief Scientific Adviser's Science Report*. Retrieved from London, UK: https://www.food.gov.uk/sites/default/files/media/document/chiefscientificadviserssciencereport%20%281%29.pdf

Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. *Science, Technology & Human Values, 30*(2), 251-290. doi:10.1177/0162243904271724

Selin, C., Rawlings, K. C., de Ridder-Vignone, K., Sadowski, J., Altamirano Allende, C., Gano, G., . . . Guston, D. H. (2017). Experiments in engagement: Designing public engagement with science and technology for capacity building. *Public Underst Sci, 26*(6), 634-649. doi:10.1177/0963662515620970

Shelley-Egan, C., & Davies, S. R. (2013). Nano-industry operationalizations of "responsibility": Charting diversity in the enactment of responsibility. *Review of Policy Research, 30*(5), 588-604. doi:10.1111/ropr.12032

Stilgoe, J., Irwin, A., & Jones, K. (2006). *The received wisdom: Opening up expert advice*. Retrieved from London, UK: https://www.demos.co.uk/files/receivedwisdom.pdf

Stilgoe, J., Lock, S. J., & Wilsdon, J. (2014). Why should we promote public engagement with science? *Public Understanding of Science, 23*(1), 4-15.

The Royal Society, British Academy, Fourniol, F., & McLaughlin, F. (2017). Data governance: public engagement review. Retrieved from London, UK: https://royalsociety.org/~/media/policy/projects/data-governance/data-governance-public-engagement-review.pdf

Torgersen, H., & Schmidt, M. (2013). Frames and comparators: How might a debate on synthetic biology evolve? *Futures, 48*, 44-54. doi:10.1016/j.futures.2013.02.002

Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and socio technical change: Imagined subjectivities of 'the public' and their implications. *Environment and Planning A, 42*(4), 931-947.

Walls, J., Pidgeon, N., Weyman, A., & Horlick-Jones, T. (2004). Critical trust: understanding lay perceptions of health and safety risk regulation. *Health, Risk & Society, 6*(2), 133-150. doi:10.1080/1369857042000219788

Warburton, D. (2009). *Evaluation of BERR's engagement of the public and other interested parties in the future of civil nuclear power in the UK: Final Report*. Retrieved from London, UK: http://www.sharedpractice.org.uk/Downloads/Nuclear_report.pdf

Wynne, B. (2006). Public engagement as a means of restoring public trust in science--hitting the notes, but missing the music? *Community Genet, 9*(3), 211-220. doi:10.1159/000092659

LINKING TEXT

Chapter 7 found that publics describe good public engagement as participation exercises that are flexible in form, focus on means and ends of data projects, empower publics, and influence policy decisions. However, participants in the government, non-technical and technical public workshops alike challenged the potential success of public engagement due to perceived public disengagement. This disengagement was underpinned by a belief that publics are generally disinterested in civic participation but also that publics did not and could not have the scientific skills needed to actively engage in government data practice. This suggests, on a broader level, that public engagement on government data use must first address mechanisms through which to address or challenge these perceptions. A feminist lens is, in particular, important for understanding and analysing these results. It would be, from a purist epistemological viewpoint, easy to accept the results as a firm truth, e.g. that because 'publics' are described as uninterested they are indeed uninterested. However, drawing on feminist research as a change-based practice, these results are instead understood as a challenge to overcome in the design and promotion of public engagement. The study results also demonstrate the way that public identity shifts and influences the perspectives of publics in deciding what good public engagement consists of. The varying publics engaged in this study discussed public engagement as activities they were familiar with, e.g. the technical publics saw themselves as mediators between government and publics. These perceptions and identifications often limit publics in imagining more inclusive and diverse forms of public engagement. Identity and perceptions were also of note in the responses of the different publics to the workshop facilitator in which technical publics were sceptical of the facilitator while non-technical publics deferred to her as more knowledgeable and powerful. Technical publics and government publics, with limitations, see themselves as more powerful and more responsible in data analysis but also in public engagement. These power relations and social assumptions challenge the ability to distribute the responsibility for government data use to non-technical

and non-governmental publics. Thus this chapter finds a key challenge for public engagement in overcoming these perceptions.

Chapter 8 also addresses Objective 3 and shifts away from a qualitative methodology to a large-scale quantitative survey. As a reminder, Objective 3 aims to examine the potential for new forms of public data engagement in government. This chapter builds on the work done in all previous chapters to develop a novel public engagement scale to help explore the nuances of public preferences for data engagement at the population level. Specifically, Chapter 8 aims to explore the relationship between different types of data, psychological ownership towards data, and public engagement preferences with data. This chapter is not currently submitted for publication. This is also the final linking text in the thesis, as Chapter 8 is the final empirical chapter. All of the results presented are synthesised and discussed in Chapter 10's Discussion.

A quantitative study was chosen for two reasons. First, to test and explore whether assumptions made in government publics about the connection between privacy and public engagement hold at a quantitative level. Second, it was an opportunity to quantitatively demonstrate from a feminist perspective, publics' interest in public engagement on government data use. As shown in Chapters 5 and 6, numbers and data are powerful communication tools with government publics. Thus, statistical rationale that demonstrates public interest was chosen as the final thesis study. The public of interest in this study is defined broadly, e.g. UK citizens over 18 years of age, in order to demonstrate interest in public engagement across a range of potential identities and characteristics. This study is seen as a final step in preparing a feminist argument for further engaging with a range of publics in the UK.

An appendix is included following this chapter with supplementary analysis not directly relevant to the question of privacy and public engagement but nevertheless important to understanding the range of preferences for public engagement among UK citizens. This is included to encourage further reflection on the nuances of public engagement. The data set derived from this study will be

released in the public domain to encourage further quantitative research on public interest in public engagement on government data use.

CHAPTER 8

Exploring public preferences for engagement with UK government data usage: A quantitative analysis

CHAPTER DECLARATION

This declaration concerns the article entitled:									
Exploring public preferences for engagement with UK government data usage: A quantitative analysis									
Publication status (tick one)									
draft manuscript	✓	**Submitted**		**In review**		**Accepted**		**Published**	
Publication details (reference)	n/a								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate predominantly executed the formulation of ideas, methodology design, work, and presentation of data. Formulation of ideas: ESR formulated the ideas of the research (90%) with supervision from JB and HD (10%). Design of methodology: ESR designed the survey (75%) with supervision from JB and HD (10%), and help in finding survey question from an RA (15%). Experimental work: ESR conducted the survey (100%). Presentation of data in journal format: ESR wrote up the results (90%) with editing and supervision from JB and HD (10%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						**Date**	18/06/2018		

8.1 INTRODUCTION

Public engagement with government data usage is a burgeoning field within public engagement with science and technology that addresses how publics can be engaged in the governance, development and usage of data and related data technologies (Michael & Lupton, 2016). Public engagement is a subset of democratic activity that focuses specifically on the inclusion of non-technical publics in the development and governance of new technologies. This traditionally encompasses a range of activities including communication, consultation, and collaboration (Rowe & Frewer, 2005) for a range of purposes including tokenism, inclusivity, and technological improvement (Fiorino, 1990; Stirling, 2004). In recent years, public engagement has shifted away from a model on communicating to publics and towards a more inclusive model of consulting and collaborating with publics (Wynne, 2006). In terms of government data usage, the aim is to ensure that publics and government share their imaginations of how data should be used and what ethical data development should look like. Publics can offer diverse and critical perspectives on the future of data technologies in society (Stilgoe, Lock, & Wilsdon, 2014; Wilsdon, Wynne, & Stilgoe, 2005).

Public awareness of and interest in how data is used by government in the United Kingdom (UK) has drastically increased in recent years (Carter, Laurie, & Dixon-Woods, 2015; Powles & Hodson, 2017). Public attention has focussed around data events like the lack of communication around the Care.Data healthcare record sharing plan (Carter et al., 2015) and more recently the use of data in political campaigning seen in the Cambridge Analytica scandal (Adams, 2018). It is increasingly relevant for government, in turn, to become more aware and interested in how publics want data to be used. Government is, of course, responsible to consider the views of the publics that it governs. Beyond a sense of democratic mandate, public engagement on data usage will become increasingly integral to decision-making as government struggles with how to decide what is and is not a good use of data technology. As described in the recent adoption of the European Parliament's General Data Protection Regulations (European

Parliament & Council of the European Union, 2016), the "processing of personal data should be designed to serve mankind." Deciding the mechanism and meaning of serving mankind through data analysis is evidently not a simple concept, and will be most usefully addressed by publicly engaged and critical debate on the future of the data society. While several government and academic researchers have explored public perceptions towards data (The Royal Society, British Academy, Fourniol, & McLaughlin, 2017), few to none have examined what publics want and why in regard to public engagement on data. This paper presents a first step in exploring public preferences for engagement around government data usage in the United Kingdom.

8.1.1 Government and data technologies in the UK

Data usage, and associated technologies, have a range of definitions. In this paper we focus on the use of data in government, particularly around data science. We define data science and related data technologies as assemblages of processes of creating, collecting, sharing, and applying information that have both technical and social purposes (Markus & Topi, 2015; McNeely & Hamh, 2014). Within government, at both national and local levels, data can be applied to administrative and policy-based processes (Joseph & Johnson, 2013). An example of administrative use is the Food Standard Agency's use of social media data and foodborne illness data to create an early warning algorithm for norovirus outbreaks (Poppy & Food Standards Agency, 2017). Similarly, road safety data has been used to examine behavioural patterns of speeding drivers to send out modified speeding notices to reduce re-offending (Graham, 2018). These kinds of administrative processes of data usage focus on using data to address the day-to-day activities and responsibilities of government services. The goal of policy-based data science is the development, evaluation, and justification of policy choices. For example, the Ministry of Justice proposed use of social media to understand public opinion of the court system (Ginnis et al., 2016). Data science can be used both as a tool to complement current practice but also for exploratory or change-oriented purposes. Data science is not only about testing solutions to a predetermined

policy area nor about addressing current administrative functions but also about developing new models of working. For example, machine learning algorithms have been used to cut down on the number of variables needed, and thus resources required, for local government to effectively predict community capacity (Piscopo, Siebes, & Hardman, 2017).

In several contexts, the application of data as described above is more theoretical than practical. In a 2015 survey of 1114 UK civil servants 62% stated that their 'organisation welcomes the use of data to inform decision making.' while 42% stated they did not 'have sufficient information to make decisions' (Civil Service World & SAS Institute Inc., 2015). While the majority of civil servants welcome data usage, less than half feel they have the necessary data to do as such. In addition, while there is a heavy focus on data in its transformative capacity for national government (Hancock, 2016; Jarrar, 2017), local government in particular struggles to realise the potential of data science (Malomo & Sena, 2017; Milakovich, 2013). While there are a range of potentially transformative data science activities, the capacity for government to realise these is limited.

8.1.2 The current state of public engagement on government data usage in the UK

Several UK government-led activities have already taken place in an attempt to understand the public view on data-related concepts including data sharing, data ethics, and machine learning, (see: Cameron, Pope, Clemence, & Ipsos MORI Social Research Institute, 2014; Davidson et al., 2013; Ginnis et al., 2016; Ipsos MORI, 2016; Sciencewise, 2012, 2014; The Royal Society et al., 2017). Much of this literature has focussed on two dimensions of perceived public concern: privacy and anonymity (Joseph & Johnson, 2013; Schintler & Kulkarni, 2014; Stough & McBride, 2014). In a 2017 review of public opinions on data usage, publics were described as having low data literacy but nevertheless concerned about how data is used. In that review and others, publics are often described as being more supportive of data use that has a clear element of public good although what that means is not often discussed (Ginnis et al., 2016; The Royal Society et al., 2017). Certain kinds of data were deemed to be of more concern than

others, for example health data. In a conjoint analysis where participants were asked to select their preferred project from government data science projects that were diverse on several different aspects of data, type of data was the largest factor in their selection (Ginnis et al., 2016, p. 48). One public engagement found that '[d]ata about who you are (i.e. personal information) is generally considered by the public to be more personal than data about what you do (i.e. behavioural data), though this distinction is likely to become increasingly spurious' (Sciencewise, 2014, p. 1). Similarly, The Royal Society et al. (2017) review discusses the confusion that publics face in defining what 'personal data' means. While not a focus of this paper, there is also a burgeoning interest in public engagement *using* data technologies, for example online e-participation and citizen science (Gagliardi et al., 2017; Lauriault & Mooney, 2014).

Due to these early exercises in determining public perceptions of data, this paper considers several dimensions of data to examine the nuances of public engagement preferences. This includes various differences in type of data as well as more general concepts of feelings of ownership. Five diverse types of data were chosen to compare in this study. These data vary on a number of dimensions including who holds that data, whether the information is created passively, whether the data is born digital, and whether it is produced in the public domain. Who holds data is included as a dimension as public engagement exercises often discuss or compare public perceptions to commercial versus government-held data, and public expectations if data held by one is shared with the other (Ipsos MORI, 2016). Whether information is created passively, e.g. automatically tracked GPS data, born digital, e.g. website data, or produced in the public domain, e.g. public posts on social media, highlight the range of differences between new and more traditional types of data and how expectations on their use may differ (Gruzd, Jacobson, Kumar, & Mai, 2017). For example, there may be different expectations about social media data that to some degree is produced and released in the public domain like Tweets versus ostensibly private information that is unknowingly produced like mobile phone GPS tracking. Data that vary on all

these dimensions are included to examine how these different dimensions of data may inform preferences for public engagement. Beyond type of data, this paper also considers how closely an individual feels that their data is owned by them. As described by Avey, Avolio, Crossley, and Luthans (2009, p. 175) feelings of ownership relate to possession that result in 'emotional, attitudinal and behavioural effects on those that experience ownership.' This study relates ownership to the concepts of privacy and personal data, in that feeling possession or ownership of data could result in or from desires for privacy of that information and stronger feelings that the data type is personal information. We include this dimension due to the strong focus on public perceptions of data research in relation to defining what is and is not personal data and private information (Sciencewise, 2014).

While public engagement exercises on data usage have so far formed a useful baseline on which to query public views on data, they do not form a full account of public preferences for future engagement on data. The limitation in any of these activities is that they are single snapshots in time of small groups of individuals about specific data concepts (Davidson et al., 2013). They do not address mechanisms of consulting and including publics in how data technologies are used on an ongoing basis. Missing in much of the literature around public engagement on government data usage is a critical exploration of how publics want to be involved and what motivates those preferences. Accordingly, this paper considers two key dimensions, type of data and ownership of data, for their connection to public engagement preferences. This paper's hypothesis is that ownership of data, and how that relates to type of data, is a key driver and mediator in public preferences for public engagement with data.

8.1.3 Aims and objectives

The aim of this study is to explore the relationship between different types of data, psychological ownership of data, and public engagement preferences with data. The first objective is to compare and contrast preferences for levels of engagement by type of data. The second objective is to use linear regression and

difference mediation testing to examine if levels of ownership mediates any relationship between type of data and preferences for levels of engagement.

8.2 MEASURES AND METHODS

8.2.1 Overview

We explore personal perceptions of data and public engagement preferences through an online survey. The survey was conducted using the hosting platform Qualtrics and the recruitment platform Prolific Academic. We aimed to recruit 1100 participants to ensure a large sample (Prolific, 2018; Qualtrics, 2018). Inclusion criteria included being over 18 years of age and currently residing in the UK. No exclusion criteria were applied. Respondents were compensated £1.68 through the Prolific platform. All participants were asked to provide their consent to release the final dataset as 'open data'. Participants first filled in demographic and other measures, then were randomised to one of five types of data. They were instructed to fill in the measures related to psychological data ownership and public engagement in reference to their type of data. All study procedures were approved by the University of Bath's Department of Psychology Research Ethics Board, reference code 18-037.

8.2.2 Main variables and measures

A range of variables were included in the survey that related to the main three variables of interest: psychological ownership of personal data, types of data, and public engagement preferences. Beyond the three key variables, various other measures were included in the survey due to their potential relevance to public engagement preferences and feelings of ownership. These variables were included for future research on the nuances of public engagement preferences but were not used in the mediation testing. A description of these variables alongside preliminary reporting on their distribution and relationship to public engagement is presented in Appendix 1: Additional Variables and Analysis at the end of this chapter. For the full survey see Appendix A: Research Tools.

8.2.3 Engagement preferences and intensity

No current scale exists that measures public engagement preferences on data usage, subsequently we designed a novel scale based on the work of Rowe and Frewer (2005). A series of seven statements were developed in relation to the range of types of engagement from communication to consultation to collaboration. Participants were asked to rank how strongly they felt about the statement from zero to ten, where ten is very strongly and zero is not at all strongly. The statements are presented in Table 8.1. Responses to these statements are presented as a summed average as well as individual scores on each statement. A weighted average was also calculated, where communication statements were counted at their absolute value, consultation statements were multiplied by two and collaboration statements were multiplied by three. This reflects that more time and resource intensive engagement activities are represented as higher numbers in the average. This was calculated as follows: Average = ((Communication 1 + Communication 2 + Communication 3) + (2 x (Consultation 1 + Consultation 2) + (3 x (Collaboration 1 + Collaboration 2))/7.

Statements were reviewed by four pilot participants for coherence prior to running the full survey. The scale was tested for internal consistency and had a high Cronbach's standardised alpha value of 0.92. All items correlated well (inter-correlation mean of 0.61), barring the seventh item which had a low correlation (ranging 0.26 to 0.35) with all other items. Dropping the seventh item increased the Cronbach's alpha to 0.95. Weighting the scale as outlined above maintained the Cronbach's alpha value at 0.92. Due to the high inter-correlation mean, and the still high alpha value with item seven maintained, all seven items are included in the final analysis. Items are included as both individual statements and as an unweighted and weighted average in ANOVA testing. In linear regression models, the weighted average is used. Participants were also asked a question related to their belief in their ability to be involved in public engagement. This was adapted from Barnett, Cooper, and Senior (2007) and the British Social Attitudes survey (Park, Curtice, Thomson, Bromley, & Phillips, 2005): 'Modern data science

is so complex that public involvement is not realistic'. Participants were asked to respond to a five-item Likert scale from strongly disagree (1) to strongly agree (5).

Table 8.1: Public engagement preference scale

Statement	Dimension of Engagement
1- 'I would want to be informed about who had access to data like this about myself'	Communication
2- 'I would want to be informed if government had access to data like this about myself'	Communication
3- 'I would want to be informed how government used data like this'	Communication
4- 'I would want to control who had access to data like this about myself'	Consultation
5- 'I would want to be asked before data like this about myself is shared with government'	Consultation
6- 'I would want to help decide how data like this, about myself and others, is used by government'	Collaboration
7- 'I would want to help government use data like this, about myself and others, better'	Collaboration

8.2.4 Psychological data ownership

A numeric scale for psychological data ownership was adapted from the Van Dyne and Pierce (2004) scale of seven-items describing personal feelings of possession originally developed for organisational psychology. The seven adapted items are: 'This is my data', 'I sense that this kind of data is OUR data', 'I feel a very high degree of personal ownership of this data', 'I sense that this is MY data', 'This is OUR data', 'Most of the people that produce this data feel as though they own the data', and 'It is hard for me to think about this data as mine'. Participants were asked to rank how strongly they felt about each of the seven statements from not at all strongly (0) to very strongly (10). The seven responses were then summed and divided by seven to calculate an average psychological ownership

towards data. The original Van Dyne and Pierce (2004) study had an internal consistency reliability Cronbach's Alpha ranging from 0.87 to 0.93 based on multiple study populations.

Table 8.2: Dimensions of type of data used in survey randomisation

	Dimension of Data				
Type of Data	**Government held?**	**Born digital?**	**Mostly passive creation?**	**In the public domain?**	**Commercially held?**
Social Media	No	Yes	No	Maybe	Maybe
GPS	No	Yes	Yes	No	Maybe
Health Records	Maybe	No	Maybe	No	No
Purchase History	No	Yes	Yes	No	Yes
Employment History	Yes	No	No	Maybe	No

8.2.5 Types of data

As outlined in the introduction, five types of data were included in the randomisation that range on a number of factors including whether government traditionally holds that type of data, whether the data is born digital, whether individuals passively or actively create that data, whether that data is typically in the public domain, and whether the data comes from commercial sources. These differences are detailed in Table 8.2. The five data types are: posts on public social media sites like Twitter, Reddit, or TripAdvisor, location (GPS) information from a mobile phone, health records from a GP, purchase history from a store card or online store, and employment history. This list is evidently not exhaustive but includes five diverse kinds of data as a starting point for examining differences in public engagement preferences.

8.2.6 Other variables and measures

Demographic variables were included to examine any differences between the five groups and subsequently if these differences would impact the validity of the randomisation. Demographic variables include age in years, employment status, gender, ethnicity, and education. Age in years is a numeric variable where age is calculated as the current year, 2018, subtracted by year of birth. Employment status is a categorical variable with eleven categories including part time work which was defined as less than 16 hours per week. Gender is a categorical variable with four categories: male, female, not listed with an option to respond in text, and prefer not to say. Ethnicity is a categorical variable with 14 categories including five broad categories: White, Mixed Heritage, Asian, Black, Other, and prefer not to say. Education is a categorical variable with nine categories.

8.2.8 Parametric and non-parametric tests of difference

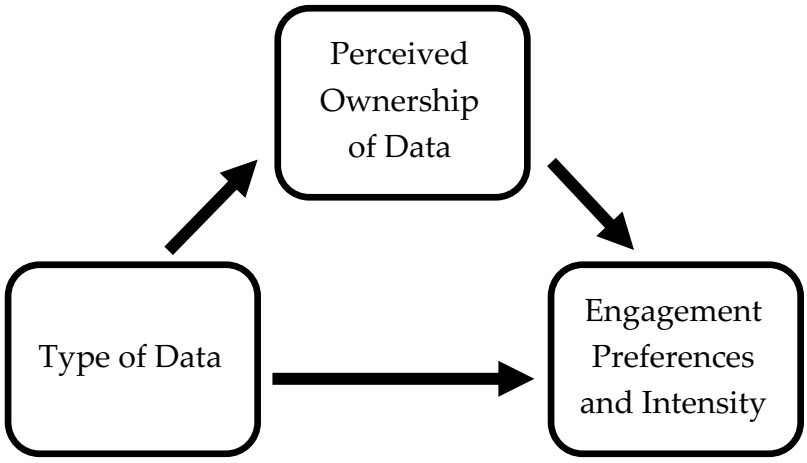
To test the validity of the randomisation, difference in distribution of demographic variables was calculated by the type of data categories. Categorical variables are reported as absolute values with percent and compared using Chi-squared tests. Numeric variables are reported as means with standard deviations and compared using one-way Analysis of Variance tests. Interval scale variables are reported as medians with the interquartile range and compared using Kruskal-Wallis tests. Statistical significance was set at $p<0.01$ for all tests. The overall distribution of each variable, not separated by type of data, is also reported. All analysis was conducted using the statistical package R (R Core Team, 2018).

8.2.9 Mediation analysis

A difference test of mediation was conducted for the effect of feelings of ownership on the relationship between type of data and desires for engagement. The proposed relationship between these variables is demonstrated in Figure 8.1. As outlined in the diagram, this mediation model tests first the relationship between type of data and engagement preferences. And second, whether perceived ownership of data partially or fully mediates that relationship. First a univariate linear regression model was calculated for the effect of type of data on

engagement and type of data on ownership. Third, a univariate linear regression model was calculated for perceived ownership on engagement preferences. For mediation to occur all of these relationships must be significant, which is set in this case at $p<0.01$. Finally, if the requirements for mediation are met a multivariate regression model is calculated for ownership and type of data on engagement (Mackinnon, Fairchild, & Fritz, 2007). Direct and indirect effects are calculated, and significance is determined at $p<0.01$ using the bootstrap method. Coefficients, standard errors and p-values are reported for all variables. All analysis was conducted using the statistical package R (R Core Team, 2018).

Figure 8.1: Direct acyclic graph for the proposed relationship between engagement preferences and intensity, perceived ownership of data and type of data.



8.3 RESULTS

8.3.1 Population description and randomisation validity

A total of 1100 participants successfully completed the survey. One participant was missing year of birth, and was subsequently not included in the statistical tests for differences in age. Demographic results and mean difference comparisons are reported in Table 8.3. Overall the sample was mostly Female (68.6%), White (91.5%), in Full-time work (46.5%) and had completed a bachelor's degree or equivalent (36%). Participants were successfully randomised to each type of data as follows: 225 in Employment History, 218 in Health Records, 234 in Location GPS data, 208 in Purchase History, and 215 in Social Media Data. The

sample did not vary by type of data significantly on any variable except average preferences for public engagement. Differences between groups in psychological ownership towards data was marginally significant at p=0.016.

8.3.2 Categorical comparisons for public engagement preferences

Table 8.4 presents the differences in mean and median value for key variables including the public engagement statements and scale, as well as psychological ownership. Overall the weighted average for public engagement preferences was 14.29 out of a possible 18.57 (Standard Deviation (S.D.) = 3.68). Participants responded highest in the Health Records category at 15.61 (S.D. = 2.91) and lowest in Employment History 13.56 (S.D. = 4.18). Median responses to individual questions were very high across all types of data ranging from a 6 to 10 out of a possible 10. The highest median response overall was for the statement 'I would want to be informed how government use data like this' at 9 (Interquartile Range (IQR) = 8-10). The lowest was for the final collaborative statement at 7 (IQR = 4-9). Participants expressed the strongest preferences for engagement in regard to the communication and collaboration statements. Differences for public engagement preferences between data types for all of the statements were significant, $p<0.001$ in all cases, as well as for the weighted ($p<0.001$) and unweighted averages ($p<0.001$).

Table 8.3: Chi-squared, Kruskal-Wallis, and ANOVA testing of any difference in demographic distribution by type of data

		Type of Data					
	All	**Employment History**	**Health Records**	**Location GPS Data**	**Purchase History**	**Social Media Data**	**p**
n	1100	225	218	234	208	215	
Age in years (mean (SD))	38.40 (11.86)	38.47 (11.28)	39.07 (11.64)	38.01 (11.68)	37.65 (12.50)	38.80 (12.28)	0.731
Employment (%)							0.338
Full time education (more than 16 hours per week)	124 (11.3)	24 (10.7)	18 (8.3)	34 (14.5)	24 (11.5)	24 (11.2)	
Part time education (less than 16 hours per week)	24 (2.2)	2 (0.9)	4 (1.8)	7 (3.0)	5 (2.4)	6 (2.8)	
Apprenticeship/Internship	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	0 (0.0)	
Working full time (more than 16 hours per week)	513 (46.6)	96 (42.7)	118 (54.1)	104 (44.4)	93 (44.7)	102 (47.4)	
Working part time (less than 16 hours per week)	141 (12.8)	38 (16.9)	24 (11.0)	30 (12.8)	22 (10.6)	27 (12.6)	
Stay at home parent	136 (12.4)	33 (14.7)	26 (11.9)	25 (10.7)	32 (15.4)	20 (9.3)	
Volunteering	13 (1.2)	4 (1.8)	2 (0.9)	2 (0.9)	1 (0.5)	4 (1.9)	
Unemployed	89 (8.1)	19 (8.4)	16 (7.3)	21 (9.0)	13 (6.2)	20 (9.3)	
Retired	52 (4.7)	9 (4.0)	7 (3.2)	10 (4.3)	14 (6.7)	12 (5.6)	
Prefer not to say	7 (0.6)	0 (0.0)	3 (1.4)	1 (0.4)	3 (1.4)	0 (0.0)	
Gender (%)							0.741
Male	342 (31.1)	75 (33.3)	69 (31.7)	68 (29.1)	60 (28.8)	70 (32.6)	
Female	755 (68.6)	150 (66.7)	149 (68.3)	165 (70.5)	147 (70.7)	144 (67.0)	
Not listed	1 (0.1)	0 (0.0)	0 (0.0)	1 (0.4)	0 (0.0)	0 (0.0)	
Prefer not to say	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.5)	
Ethnicity (%)							0.628
White	1007 (91.5)	204 (90.7)	196 (89.9)	216 (92.3)	193 (92.8)	198 (92.1)	
Mixed Heritage - White and Black African	3 (0.3)	2 (0.9)	0 (0.0)	1 (0.4)	0 (0.0)	0 (0.0)	
Mixed Heritage - Other	6 (0.5)	3 (1.3)	1 (0.5)	1 (0.4)	1 (0.5)	0 (0.0)	
Asian - Indian	12 (1.1)	3 (1.3)	4 (1.8)	2 (0.9)	2 (1.0)	1 (0.5)	
Asian - Bangladeshi	4 (0.4)	1 (0.4)	0 (0.0)	2 (0.9)	0 (0.0)	1 (0.5)	
Mixed Heritage - White and Black Caribbean	7 (0.6)	0 (0.0)	3 (1.4)	1 (0.4)	2 (1.0)	1 (0.5)	
Mixed Heritage - White and Asian	5 (0.5)	1 (0.4)	2 (0.9)	1 (0.4)	0 (0.0)	1 (0.5)	
Asian - Chinese	13 (1.2)	2 (0.9)	5 (2.3)	1 (0.4)	2 (1.0)	3 (1.4)	
Asian - Pakistani	11 (1.0)	2 (0.9)	3 (1.4)	2 (0.9)	1 (0.5)	3 (1.4)	
Asian - Other	6 (0.5)	2 (0.9)	0 (0.0)	1 (0.4)	2 (1.0)	1 (0.5)	
Black - Caribbean	4 (0.4)	1 (0.4)	0 (0.0)	2 (0.9)	0 (0.0)	1 (0.5)	
Black - African	10 (0.9)	4 (1.8)	3 (1.4)	1 (0.4)	1 (0.5)	1 (0.5)	
Black - Other	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	
Any other ethnic background	4 (0.4)	0 (0.0)	0 (0.0)	2 (0.9)	0 (0.0)	2 (0.9)	
Prefer not to say	7 (0.6)	0 (0.0)	1 (0.5)	1 (0.4)	4 (1.9)	1 (0.5)	

Table 8.3 ***cont'd*****:** Chi-squared, Kruskal-Wallis, and ANOVA testing of any difference in demographic distribution by type of data

		Type of Data					
	All	**Employment History**	**Health Records**	**Location GPS Data**	**Purchase History**	**Social Media Data**	**p**
Education (%)							0.206
GCSE/O-Level/CSE/Scottish Standards or Equivalent	196 (17.8)	48 (21.3)	42 (19.3)	34 (14.5)	34 (16.3)	38 (17.7)	
Vocational qualifications	98 (8.9)	21 (9.3)	21 (9.6)	11 (4.7)	24 (11.5)	21 (9.8)	
A-level/Scottish Highers or Equivalent	224 (20.4)	33 (14.7)	52 (23.9)	57 (24.4)	41 (19.7)	41 (19.1)	
Bachelor's Degree or Equivalent	396 (36.0)	86 (38.2)	61 (28.0)	95 (40.6)	79 (38.0)	75 (34.9)	
Masters/PhD or Equivalent	154 (14.0)	31 (13.8)	37 (17.0)	32 (13.7)	25 (12.0)	29 (13.5)	
Other	13 (1.2)	3 (1.3)	2 (0.9)	2 (0.9)	3 (1.4)	3 (1.4)	
No formal qualifications	17 (1.5)	2 (0.9)	3 (1.4)	3 (1.3)	2 (1.0)	7 (3.3)	
Don't know	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Prefer not to say	2 (0.2)	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	

Table 8.4: Chi-squared, Kruskal-Wallis, and ANOVA testing of any difference in distribution of public engagement preferences and psychological ownership by type of data

		Type of Data					
	All	**Employment History**	**Health Records**	**Location GPS Data**	**Purchase History**	**Social Media Data**	**p**
n	1100	225	218	234	208	215	
Nominal scale (1-5) agreeance to "Modern data science is so complex that public involvement is not realistic." (median [IQR])	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	0.81
Average psychological ownership of data (mean (SD))	5.53 (1.37)	5.56 (1.38)	5.79 (1.24)	5.37 (1.40)	5.46 (1.28)	5.48 (1.48)	0.016
Average desires for public engagement with data (mean (SD))	7.90 (2.00)	7.44 (2.27)	8.61 (1.52)	8.11 (1.77)	7.65 (2.05)	7.70 (2.10)	<0.001
Weighted average desires for public engagement with data (mean (SD))	14.29 (3.68)	13.56 (4.18)	15.61 (2.91)	14.52 (3.27)	13.74 (3.76)	14.00 (3.81)	<0.001
Public Engagement Scale Questions (median [IQR])							
I would want to be informed about who had access to data like this about myself.	9.00 [7.00, 10.00]	8.00 [7.00, 10.00]	10.00 [8.00, 10.00]	10.00 [8.00, 10.00]	9.00 [7.00, 10.00]	8.00 [7.00, 10.00]	<0.001
I would want to be informed if government had access to data like this about myself.	9.00 [7.00, 10.00]	8.00 [6.00, 10.00]	10.00 [8.00, 10.00]	10.00 [8.00, 10.00]	9.00 [7.00, 10.00]	9.00 [7.00, 10.00]	<0.001
I would want to be informed how government used data like this.	9.00 [8.00, 10.00]	9.00 [7.00, 10.00]	10.00 [8.00, 10.00]	10.00 [8.00, 10.00]	9.00 [7.00, 10.00]	9.00 [7.00, 10.00]	<0.001
I would want to control who had access to data like this about myself.	9.00 [7.00, 10.00]	8.00 [6.00, 10.00]	10.00 [8.00, 10.00]	10.00 [8.00, 10.00]	9.00 [7.00, 10.00]	9.00 [7.00, 10.00]	<0.001
I would want to be asked before data like this about myself is shared with government.	9.00 [7.00, 10.00]	8.00 [6.00, 10.00]	10.00 [9.00, 10.00]	10.00 [8.00, 10.00]	9.00 [7.00, 10.00]	9.00 [7.00, 10.00]	<0.001
I would want to help decide how data like this, about myself and others, is used by government.	9.00 [7.00, 10.00]	8.00 [6.00, 10.00]	10.00 [8.00, 10.00]	9.00 [7.00, 10.00]	8.00 [6.00, 10.00]	9.00 [7.00, 10.00]	<0.001
I would want to help government use data like this, about myself and others, better.	7.00 [4.00, 9.00]	7.00 [5.00, 8.00]	8.00 [5.00, 10.00]	6.00 [4.00, 8.75]	6.00 [3.00, 8.00]	6.00 [4.00, 9.00]	<0.001

Table 8.5: Mediation difference testing using simple linear regression*

Model	Univariate Coefficient	Standard Error	P-Value
Type of Data on Psychological Ownership Towards Data			
Employment Data	Ref.	Ref.	Ref.
Social Media Data	-0.06228	0.12921	0.629885
Store Purchase History	-0.06606	0.13063	0.613147
Health Records	0.22307	0.12875	0.083454
Location/GPS Data	-0.17535	0.12646	0.16584
Type of Data on Public Engagement Preferences			
Employment Data	Ref.	Ref.	Ref.
Social Media Data	0.4374	0.3444	0.2044
Store Purchase History	0.1743	0.3474	0.61594
Health Records	2.0455	0.3432	3.40E-09
Location/GPS Data	0.9589	0.3372	0.00454
Psychological Ownership Towards Data on Public Engagement Preferences			
Psychological ownership towards data	0.86675	0.07689	<2e-16

*Mediation was not undertaken as type of data had no relationship with ownership.

8.3.3 Mediation of public engagement preferences and type of data

Table 8.5 presents the results from the mediation analysis. The relationship of public engagement preferences regressed on ownership was highly significant ($\beta = 0.87$, p=<2e-16). Such that for every one-point increase in the ownership scale, the public engagement preference scale increased by 0.87 points on average. In other words, a greater feeling of ownership indicated a stronger preference for public engagement. The relationship of public engagement preferences regressed on type of data where employment data was the reference category was significant in two cases. Health records (2.05, p=3.40e-09) and location/GPS data (0.96, p=0.0045) were significantly different to employment data in preferences for public engagement. This indicates that participants expressed stronger desires for public engagement when randomised to the health records and location/GPS data categories. Health records had the strongest association, participants in this group on average responded 2.05 points higher on the public engagement preference scale compared to those in the employment data group. The relationship of data

ownership regressed on type of data was not significant in any case. Psychological ownership towards data thus did not meet the requirements as a mediating variable.

8.4 DISCUSSION

8.4.1 Summary of results and comparisons to previous research

This study aimed to explore public preferences for engagement on government data usage. Overall, the results demonstrate that participants in this study expressed interest in understanding and negotiating how their data is used by government. They were most interested in the dimensions of communication and consultation, as shown in the higher mean values (median 9 out of 10 where 10 is a very strong preference to be communicated or consulted) in their response to those statements on the public engagement preferences scale. They were less but still highly interested in deliberative and collaborative activities to help government decide how to use data better (median response 7 out of 10). As this scale is novel to this study there are evidently no other studies to directly compare these results, however there is a history of scholarship on how best to involve publics that can be drawn upon to reflect on these results (Rowe & Frewer, 2005; Stilgoe et al., 2014). A common theme in this literature is demonstrating and examining the interest of publics in scientific governance. Several authors discuss how publics are imagined by policy-makers to be difficult, irrational, and at worst uninterested (Burningham, Barnett, & Walker, 2014; Laurent, 2007; Shelley-Egan & Davies, 2013). Subsequently, authors contend that while not all publics are equally interested in scientific governance, apparent disinterest often masks a sense of fatalism on whether publics can have an impact on technological development (Groves, 2011; Stilgoe, Irwin, & Jones, 2006; Willis & Wilsdon, n.d.). In other words publics wish to be involved but the mechanisms and systems of technological governance are opaque and consequently difficult to influence.

This paper provides evidence that publics within the UK may indeed be interested in governance in the context of government data use. The slightly lower interest in collaborative engagement may reflect the time commitment needed to

be involved in these kinds of activities or indeed the aforementioned feelings of fatalism or the ability to engage in more technical activities. Michael and Lupton (2016) contend that due to the complexity of concepts like big data, publics will face challenges in engaging with these technologies if they are not highly digitally literate. As evidenced in Table 8.4, participant response was mixed (median response 3 out of 5) on how much they agreed with the statement "Modern data science is so complex that public involvement is not realistic." These results suggest that better consideration is needed of how public engagement could build capacity and confidence to participate in data technology governance (Selin et al., 2017). Community data initiatives like those described by Taylor et al. (2015) where data collection and thus data literacy is embedded in small-scale community projects is one such example of mixing public engagement and capacity building. Overall, the study participants evidenced a high interest in what is done with data about themselves by government across a range of potential public engagement activities including communication, consultation, and collaboration.

8.4.2 Reflecting on the contingencies of public engagement preferences

Psychological ownership and type of data were both significantly associated with a higher preference for public engagement. Participants felt most strongly about wanting to be engaged around health records and GPS data compared to the reference category of employment data. As discussed earlier, type of data is often considered as a key factor in how publics respond to government data projects (Ginnis et al., 2016). And the results found here echo a common theme in data engagement that health data is of particular interest (Aitken, de St Jorre, Pagliari, Jepson, & Cunningham-Burley, 2016). Health records and GPS data express a range of differences and similarities as set out in the methods. Health records may be held by government while GPS information is not, as well GPS data is generally born digital while health data is not. However, both types of data are not typically in the public domain and both are generally passively created in that individuals are not actively producing this information like social media data.

This suggests that data that is generally created passively and not publicly available may require greater consultation and engagement of publics. Interestingly, purchase history was not as strongly associated with a higher preference for public engagement despite it also being passively created. There are likely further nuances to these kinds of data, and how those nuances contribute to public engagement preferences, that require further research.

Participants who felt a stronger sense of ownership of data felt more strongly about wanting engagement on data usage. However, sense of ownership was not statistically associated with type of data. Therefore, the relationship between type of data and engagement preferences was not mediated by psychological ownership. Thus, we did not find support for our hypothesis that feelings of ownership causally explained this relationship. Ownership and type of data were important but separate dimensions in understanding variations in public engagement preferences. Individuals who expressed a strong feeling of ownership towards their data did so regardless of type of data. These results challenge the conclusions from a Sciencewise (2014) consultation that hypothesised that more personal data, e.g. data that was associated with stronger feelings of ownership, would include information about who you are versus what you do. In fact, this study demonstrates that participants expressed no variation in views of ownership by type of data.

These results challenge the accepted paradigm within literature around government data use, i.e. that publics are most concerned about keeping information like health records private and are subsequently a potential road block to using sensitive data more widely (Ginnis et al., 2016; Joseph & Johnson, 2013). While participants do want to be consulted, and separately participants may have inclinations towards privacy, these results do not support the viewpoint that privacy and consent seeking are associated with type of data. Although health data was perceived by participants to be important to be consulted on, this was not related to a sense of ownership and privacy. In other words, participants did not necessarily want to make that information private in general rather

participants expressed a strong desire for influence in understanding and debating how that data is used. Simply, these results suggest that participants want greater influence in deciding what good government data use looks like, particularly influence that moves beyond simple models of consent.

8.4.3 Strengths and limitations

This the first empirical study to explore public engagement preferences on government data use. We also add to a growing literature around understanding publics' views on government data use. This is the first study to test quantitatively that type of data and perceptions of ownership of data are important to how publics wish to be engaged. This work is preliminary in nature and further qualitative and quantitative research is needed to evaluate these results. Particularly, the public engagement preferences scale, while based on a review of public engagement with science and technology, is not comprehensive. It does not include other forms of engagement like capacity building or specific forms of engagement like citizen forums. It also does not include non-normative political behaviours nor self-organised public behaviours that publics may use to express concern over government use of data. As well the high mean values and narrow standard deviation of the public engagement scale could suggest a potential response bias, in that participants reported higher desires for engagement than they may actually hold. This is evidently a risk in any online survey, and a large sample size was used in order to minimize this possibility. Finally, as the scales and measures used in this study were proxy measures for complex psychological concepts like privacy, ownership, and perceptions of data, the results may not comprehensively represent public preferences for public engagement on government data use. Further work is needed to explore the validity of the approach taken here and whether other scales and measures may better represent these concepts.

8.5 CONCLUSION

This study explores public preferences for engagement on government data usage in the UK. We found that participants expressed a strong interest in a range

of public engagement activities including communication on data usage, consultation on what their data can be used for, and collaboration on helping government use data better. We also add to a growing literature that demonstrates that type of data, particularly health data, and feelings of data ownership are important for predicting the level of engagement that publics prefer. However, we did not find support for our hypothesis that feelings of ownership mediated the relationship between type of data and engagement preferences. These results challenge the view that publics have a strong desire to keep certain kinds of data private. On a broader level, the results of this study support that publics have a high expectation of government to be more transparent in how they use data but also more open to public influence in debating the future of data use in government.

APPENDIX 1: ADDITIONAL VARIABLES AND ANALYSIS

Description of additional survey variables

Several additional variables were included due to their hypothesised relevancy to public engagement preferences or feelings of ownership toward data. They also enable further study, beyond this paper, that examines how public engagement preferences vary or indeed correlative associations between a range of political and public variables in the context of data governance. They are described in detail below for future reference. The theoretical rationale for inclusion is as follows. Personal and public efficacy, as well as measures of digital skills, were included due to the role of capacity in science engagement, i.e. the belief in the ability to participate (Selin et al., 2017). Trust in others and efficacy were also included as they are predictors of political participation (Anderson, 2010). Personal political behaviours and views of citizenship were included due to the connection between public engagement and the governance of science and technology (Braun & Konninger, 2017).

Perceptions of personal and public efficacy and trust in relation to government

Personal efficacy in relation to government was measured by four items. First, participants were asked how much they felt they know about how local government works as a four-item Likert scale from nothing at all (1) to a great deal (4) (Fox, Blackwell , Fowler, Mackay, & Boga Mitchell, 2017). Second, how much influence they feel they have over decision-making in their local area (Fox et al., 2017), as a four-item Likert scale from nothing at all (1) to a great deal (4). Third adapted from Anderson (2010), how much they feel local officials care about what they think is important for their local area as a four-item Likert scale from not at all (1) to very much (4). Fourth, how confident they are in their own ability to participate in local politics as a six-item Likert scale from don't know (1) to completely confident (6) (The European Social Survey, 2017).

Public efficacy was measured by asking participants how much they agree with the following statement 'When people like me get involved in politics they

really can change the way their local area is run' as a five-item Likert scale from strongly disagree (1) to strongly agree (5) (Fox et al., 2017). Participants were asked one question related to how satisfied they are with the way local government is doing its job as a four-item Likert scale from very dissatisfied (1) to very satisfied (4), adapted for local government from the The European Social Survey (2017). Participants were also asked how much they feel others can be trusted, as found in the New Economics Foundation (Michaelson, Mahony, & Schifferes, 2012), as a Likert scale from zero ('Can't be too careful') to ten ('Most people can be trusted').

Political behaviours

Three questions measured political behaviours. The first two are categorical response variables adapted from the Hansard Society's (Fox et al., 2017) Audit of Political Engagement 14 asking whether participants voted in the 2017 General Election and the 2016 EU Referendum. These two variables are reported as categorical with four categories: Yes, No, Don't Know, or Wasn't Eligible. The third question is a political behaviours scale also from the Hansard Society's (Fox et al., 2017) Audit of Political Engagement 14 with fourteen items measuring a range of normative political behaviours asking respondents whether they had done any of the following to influence decisions, laws or policies in the last 12 months. One additional item not included in the Hansard Scale was added asking whether participants had submitted a freedom of information request. Responses included No, Maybe, and Yes. Responses were summed across the fourteen dimensions where Yes was 2, Maybe was 1 and No was 0. A numeric average was calculated from the total summed scores divided by fourteen.

Perceptions of citizenship and interest in politics

Interest in politics was measured by two questions adapted from the Hansard Society's (Fox et al., 2017) Audit of Political Engagement 14 asking participants how interested they are in national and local politics both as a four-item Likert scale from not at all interested (1) to very interested (4). Perceptions of citizenship were measured with the Civic Participation Scale designed by Haste

and Hogan (2006) with ten items related to normative citizenship behaviours. Participants ranked how important they felt each item was in being a good citizen on a five-item Likert scale from very unimportant (1) to very important (5). The responses on each item were summed and divided by ten to calculate a numeric average for perceptions of good citizenship.

Digital skills

Digital skills were measured by two questions that indicate higher or lower level digital skills. First lower-level skills were measured by the Oxford Internet Institute's Social Digital Skills scale (van Deursen, Helsper, & Eynon, 2014) with five statements about online social behaviour. Responses are reported as a numeric average of answers to a six-item Likert scale from 'I don't understand what you mean by that' (1) to 'very true of me' (5). Higher-level skills are measured by an adapted question from the Europass (2018) Digital Competences Self-assessment grid. Participants were asked to select one of four statements that best matched their skills in programming competence. Programming skills is a categorical variable with four categories. The wording of two of the competences were drawn from the grid, namely "I know the basics of one programming language" and replacing the term 'several' with 'one or more', "I can use one or more programming languages to design, create and modify databases with a computer tool." Two types of competences were added to ensure all competences of participants were expressed, this included "I don't know what programming is" and "I am aware of what programming is but cannot use a programming language".

Preliminary analysis

Tests for differences in distribution between the five study groups was conducted. Categorical variables are reported as absolute values with percent and compared using Chi-squared tests. Numeric variables are reported as means with standard deviations and compared using one-way Analysis of Variance tests. Interval scale variables are reported as medians with the interquartile range and compared using Kruskal-Wallis tests. Statistical significance was set at $p<0.01$ for

all tests. Univariate and multivariate models were also calculated using multiple and simple linear regression for public engagement preferences regressed on all other variables, *not* separated by type of data. Variables that met significance at $p<0.01$ were included in a multivariate regression model. Coefficients, standard errors and p-values are reported for all variables. The adjusted coefficient of determination/R-squared is reported for all univariate and multivariate models. All analysis was conducted using the statistical package R (R Core Team, 2018).

Preliminary results modelling public engagement preferences

39 participants did not fill in the question on confidence to participate in local politics due to an error in the online survey software. This question was subsequently dropped from any analysis. One participant was missing age and was dropped in multivariate and univariate regressions that included age. Table A1 presents preliminary results from statistical tests to examine if there are any differences in the distribution of covariates by study group. No differences were found, however the Kruskal-Wallis test for trust in others neared significance at $p=0.02$. Table A2 presents exploratory analysis for univariate linear regression for public engagement preferences regressed on all variables: gender, age, voting in the EU referendum, interest in local politics, self-reported digital skills, and beliefs in good citizenship were significant at $p<0.01$. Of these variables, beliefs in good citizenship had the largest influence on the adjusted R-squared at $R^2=0.025$. A multivariate model including all of these variables, as well as psychological ownership towards data and type of data was calculated. The adjusted R-squared of the multivariate model was 0.1614.

Preliminary discussion on nuances of public engagement

Building a predictive model for public engagement was not the main focus of the paper but nevertheless we present preliminary results on how other variables are associated with public engagement in the context of data usage. Of particular note, some measures of political behaviours and beliefs were associated with public engagement preferences and some were not. Stronger beliefs in what

behaviours made a good citizen was most associated with the outcome of public engagement preferences. Whereas local government satisfaction and feelings of personal capacity to engage in politics were not associated with public engagement preferences. This could represent natural variation in what these scales measure, however it could also reflect that scientific governance, i.e. public engagement, is not directly understood in the same way as civic participation. Thus, asking individuals about their voting behaviour may not indicate the same kind of behaviour as wanting to be consulted about how their data is used. Data literacy, i.e. digital skills, was associated with public engagement preferences which could support the argument that capacity building should be a key part of public engagement on data usage (Selin et al., 2017). However, taken together these concepts did not highly explain public engagement preferences as demonstrated by the relatively low coefficient of determination in the overall model (R^2=0.1614), thus other variables not measured are likely important to this outcome.

Table A1: Chi-squared, Kruskal-Wallis, and ANOVA testing of any difference in covariate distribution by type of data

		Type of Data					
	All	**Employment History**	**Health Records**	**Location GPS Data**	**Purchase History**	**Social Media Data**	**p**
Did you vote in the 2017 General Election? (%)							0.247
Yes	912 (82.9)	178 (79.1)	184 (84.4)	194 (82.9)	182 (87.5)	174 (80.9)	
No	135 (12.3)	32 (14.2)	28 (12.8)	25 (10.7)	19 (9.1)	31 (14.4)	
Don't Know	7 (0.6)	2 (0.9)	2 (0.9)	1 (0.4)	2 (1.0)	0 (0.0)	
Wasn't Eligible	46 (4.2)	13 (5.8)	4 (1.8)	14 (6.0)	5 (2.4)	10 (4.7)	
Did you vote in the 2016 EU Referendum? (%)							0.169
Yes	888 (80.7)	180 (80.0)	175 (80.3)	190 (81.2)	174 (83.7)	169 (78.6)	
No	143 (13.0)	28 (12.4)	35 (16.1)	22 (9.4)	26 (12.5)	32 (14.9)	
Don't Know	7 (0.6)	3 (1.3)	1 (0.5)	2 (0.9)	1 (0.5)	0 (0.0)	
Wasn't Eligible	62 (5.6)	14 (6.2)	7 (3.2)	20 (8.5)	7 (3.4)	14 (6.5)	
Nominal scale (1-4) "How much, if anything, do you feel you know about how local government works?" (median [IQR])	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	0.996
Nominal scale (1-4) "How much influence, if any, do you feel you have over decision making in your local area" (median [IQR])	3.00 [3.00, 4.00]	3.00 [3.00, 3.00]	3.00 [3.00, 4.00]	3.00 [3.00, 4.00]	3.00 [3.00, 3.25]	3.00 [3.00, 3.00]	0.81
Nominal scale (1-4) "How much do you believe that local officials care about what you think is important for your local area" (median [IQR])	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	0.234
Nominal scale of agreeance (1-5) to "When people like me get involved in politics, they really can change the way their local area is run" (median [IQR])	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]	0.296
Nominal scale (1-4) of satisfaction with the way local government in their area is doing its job (median [IQR])	2.00 [2.00, 3.00]	2.00 [2.00, 3.00]	2.00 [2.00, 3.00]	2.00 [2.00, 3.00]	3.00 [2.00, 3.00]	2.00 [2.00, 3.00]	0.881
Nominal scale interest in national politics (1-4) (median [IQR])	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	0.974
Nominal scale interest in local politics (1-4) (median [IQR])	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	2.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	0.835
Average beliefs in good citizenship scale (mean (SD))	1.91 (0.35)	1.87 (0.34)	1.91 (0.34)	1.91 (0.36)	1.92 (0.35)	1.92 (0.37)	0.603
Average political behaviours scale (mean (SD))	1.45 (0.35)	1.45 (0.35)	1.47 (0.35)	1.48 (0.36)	1.42 (0.33)	1.45 (0.35)	0.426
Median trust in others (median [IQR])	5.00 [3.00, 7.00]	6.00 [3.00, 7.00]	6.00 [4.00, 7.00]	5.00 [3.00, 7.00]	5.00 [3.00, 7.00]	5.00 [3.00, 7.00]	0.02
Average self-reported digital social skills (mean (SD))	4.48 (0.76)	4.48 (0.73)	4.50 (0.67)	4.48 (0.75)	4.48 (0.74)	4.44 (0.90)	0.918
Competence in computer programming (%)							0.418
I don't know what programming is	70 (6.4)	16 (7.1)	15 (6.9)	10 (4.3)	17 (8.2)	12 (5.6)	
I am aware of what programming is but cannot use a programming language	688 (62.5)	148 (65.8)	137 (62.8)	143 (61.1)	130 (62.5)	130 (60.5)	
I know the basics of one programming language.	239 (21.7)	43 (19.1)	52 (23.9)	57 (24.4)	36 (17.3)	51 (23.7)	
I can use one or more programming languages to design, create and modify databases with a computer tool.	103 (9.4)	18 (8.0)	14 (6.4)	24 (10.3)	25 (12.0)	22 (10.2)	

Table A2: Univariate and multivariate linear regression for covariates on preferences for public engagement with data

	Univariate				Multivariate			
Variable	**Coefficient**	**Standard Error**	**P-Value**	**Adjusted R^2**	**Coefficient**	**Standard Error**	**P-Value**	**Adjusted R^2**
								0.1614
Type of Data								
Employment Data	Ref.	Ref.	Ref.	0.0363	Ref.	Ref.	Ref.	
Social Media Data	0.4374	0.3444	0.2044		0.529034	0.322402	0.101104	
Store Purchase History	0.1743	0.3474	0.61594		0.281897	0.325	0.385928	
Health Records	2.0455	0.3432	3.40E-09		1.838957	0.321348	1.36E-08	
Location/GPS Data	0.9589	0.3372	0.00454		1.116484	0.315666	0.000422	
Psychological ownership towards data	0.86675	0.07689	<2e-16	0.1029	0.783921	0.075865	< 2e-16	
Interest in Local Politics	0.591	0.142	0.000	0.015	0.263544	0.140361	0.060701	
Average self-reported digital social skills	0.576	0.145	0.000	0.013	0.420168	0.13619	0.002086	
Average beliefs in good citizenship scale	-1.676	0.311	0.000	0.025	-0.916913	0.30865	0.003036	
Age in years	0.031	0.009	0.001	0.009	0.017577	0.008819	0.046498	
Gender								
Male	ref.	ref.	ref.	0.003	ref.	ref.	ref.	
Female	0.618	0.239	0.010		0.445012	0.220487	0.043804	
Not listed	0.993	3.678	0.787		2.367067	3.390224	0.485199	
Prefer not to say	-0.007	2.605	0.998		1.407814	2.396408	0.557011	
Employment								
Full time education (more than 16 hours per week)	ref.	ref.	ref.	0.001	.	.	.	
Part time education (less than 16 hours per we	-0.043	0.820	0.959		.	.	.	
Apprenticeship/Internship	2.993	3.693	0.418		.	.	.	
Working full time (more than 16 hours per week)	0.383	0.368	0.298		.	.	.	
Working part time (less than 16 hours per week)	0.374	0.453	0.409		.	.	.	
Stay at home parent	0.865	0.457	0.059		.	.	.	
Volunteering	1.773	1.072	0.098		.	.	.	
Unemployed	0.508	0.511	0.320		.	.	.	
Retired	0.828	0.608	0.173		.	.	.	
Prefer not to say	-1.823	1.429	0.202		.	.	.	

Table A2 *cont'd*: **Univariate and multivariate linear regression for covariates on preferences for public engagement with data**

	Univariate				Multivariate			
Variable	**Coefficient**	**Standard Error**	**P-Value**	**Adjusted R^2**	**Coefficient**	**Standard Error**	**P-Value**	**Adjusted R^2**
Ethnicity								
White	ref.	ref.	ref.	0.006				
Mixed Heritage - White and Black Caribbean	1.423	1.391	0.307		.	.	.	
Mixed Heritage - White and Black African	1.702	2.121	0.423		.	.	.	
Mixed Heritage - White and Asian	-0.993	1.645	0.546		.	.	.	
Mixed Heritage - Other	-4.370	1.502	0.004		.	.	.	
Asian - Chinese	0.596	1.024	0.561		.	.	.	
Asian - Indian	1.226	1.065	0.250		.	.	.	
Asian - Pakistani	-1.134	1.112	0.308		.	.	.	
Asian - Bangladeshi	1.250	1.838	0.497		.	.	.	
Asian - Other	0.773	1.502	0.607		.	.	.	
Black - Caribbean	3.642	1.838	0.048		.	.	.	
Black - African	0.892	1.166	0.444		.	.	.	
Black - Other	3.035	3.670	0.409		.	.	.	
Any other ethnic background	1.464	1.838	0.426		.	.	.	
Prefer not to say	0.933	1.391	0.503		.	.	.	
Education								
GCSE/O-Level/CSE/Scottish Standards or Equivalent	ref.	ref.	ref.	0.003	.	.	.	
Vocational qualifications	-0.162	0.454	0.722		.	.	.	
A-level/Scottish Highers or Equivalent	-0.045	0.359	0.900		.	.	.	
Bachelor Degree or Equivalent	-0.611	0.321	0.057		.	.	.	
Masters/PhD or Equivalent	0.197	0.396	0.619		.	.	.	
Other	-0.570	1.052	0.588		.	.	.	
No formal qualifications	1.280	0.929	0.168		.	.	.	
Don't Know	na.	na.	na.		.	.	.	
Prefer not to say	-0.278	2.610	0.915		.	.	.	

Table A2 *cont'd*: Univariate and multivariate linear regression for covariates on preferences for public engagement with data

Variable	Univariate Coefficient	Standard Error	P-Value	Adjusted R^2	Multivariate Coefficient	Standard Error	P-Value	Adjusted R^2
Did you vote in the 2017 General Election?								
Yes	ref.	ref.	ref.	0.015	.	.	.	
No	-0.724	0.337	0.032					
Don't Know	-5.516	1.385	0.000		.	.	.	
Wasn't Eligible	-0.016	0.552	0.977		.	.	.	
Did you vote in the 2016 EU Referendum?								
Yes	ref.	ref.	ref.	0.006	.	.	.	
No	-0.897	0.331	0.007		.	.	.	
Don't Know	-2.047	1.392	0.142		.	.	.	
Wasn't Eligible	-0.280	0.482	0.561		.	.	.	
"How much, if anything, do you feel you know about how local government works?"	-0.159	0.432	0.712	-0.001	.	.	.	
"How much influence, if any, do you feel you have over decision making in your local area"	0.091	0.252	0.717	-0.001	.	.	.	
"How much do you believe that local officials care about what you think is important for your local area"	0.074	0.267	0.780	-0.001	.	.	.	
"When people like me get involved in politics, they really can change the way their local area is run"	-0.168	0.186	0.368	0.000	.	.	.	
"How satisfied are you with the way local government in your area is doing its job?"	0.495	0.331	0.135	0.001	.	.	.	
"Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?"	-0.024	0.051	0.641	-0.001	.	.	.	
Interest in National Politics	0.258	0.129	0.045	0.003				
Average political behaviours scale	0.561	0.319	0.079	0.002				
Please select the relevant statement which you feel applies to your competence in computer programming								
I don't know what programming is	ref.	ref.	ref.	-0.002	.	.	.	
I am aware of what programming is but cannot use a programming language	-0.145	0.462	0.754		.	.	.	
I know the basics of one programming language.	-0.064	0.501	0.898		.	.	.	
I can use one or more programming languages to design, create and modify databases with a computer tool.	-0.287	0.571	0.615		.	.	.	
"Modern data science is so complex that public involvement is not realistic."	-0.007	0.194	0.973	-0.001	.	.	.	

REFERENCES

Adams, T. (2018). Facebook's week of shame: the Cambridge Analytica fallout. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence

Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016). Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics, 17*(1), 73. doi:10.1186/s12910-016-0153-x

Anderson, M. R. (2010). Community psychology, political efficacy, and trust. *Political Psychology, 31*(1), 59-84. doi:10.1111/j.1467-9221.2009.00734.x

Avey, J. B., Avolio, B. J., Crossley, C. D., & Luthans, F. (2009). Psychological ownership: theoretical extensions, measurement and relation to work outcomes. *Journal of Organizational Behavior, 30*(2), 173-191. doi:10.1002/job.583

Barnett, J., Cooper, H., & Senior, V. (2007). Belief in public efficacy, trust, and attitudes toward modern genetic science. *Risk Analysis, 27*(4), 921-933. doi:10.1111/j.1539-6924.2007.00932.x

Braun, K., & Konninger, S. (2017). From experiments to ecosystems? Reviewing public participation, scientific governance and the systemic turn. *Public Underst Sci*, 963662517717375. doi:10.1177/0963662517717375

Burningham, K., Barnett, J., & Walker, G. (2014). An Array of Deficits: Unpacking NIMBY Discourses in Wind Energy Developers' Conceptualizations of Their Local Opponents. *Society & Natural Resources, 28*(3), 246-260. doi:10.1080/08941920.2014.933923

Cameron, D., Pope, S., Clemence, M., & Ipsos MORI Social Research Institute. (2014). *Dialogue on Data - Exploring the public's views on using linked administrative data for research purposes*. Retrieved from London, UK: https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics, 41*(5), 404-409. doi:10.1136/medethics-2014-102374

Civil Service World, & SAS Institute Inc. (2015). *Using big data to support a new government: Research survey report*. Retrieved from https://www.sas.com/content/dam/SAS/en_gb/doc/whitepaper1/PublicSector/CSW-SAS-research-paper.pdf

Davidson, S., McLean, C., Treanor, S., Ipsos MORI, Aitken, M., Cunningham-Burley, S., . . . University of Edinburgh. (2013). *Public acceptability of data sharing between the public, private and third sectors for research purposes*. Retrieved from

Europass. (2018). Digital Competence. Retrieved from https://europass.cedefop.europa.eu/resources/digital-competences

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), (2016).

Fiorino, D. J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. *Science Technology Human Values, 15*(2), 226-243.

Fox, R., Blackwell , J., Fowler, B., Mackay, L., & Boga Mitchell, L. (2017). *Audit of Political Engagement 14: The 2017 Report*. Retrieved from London, UK: https://assets.ctfassets.net/xkbace0jm9pp/1vNBTsOEiYciKEAqWAmEKi/c9cc36b98f60328c0327e313ab37ae0c/Audit_of_political_Engagement_14__2017_.pdf

Gagliardi, D., Schina, L., Sarcinella, M. L., Mangialardi, G., Niglia, F., & Corallo, A. (2017). Information and communication technologies and public participation: interactive maps and value added for citizens. *Government Information Quarterly, 34*(1), 153-166. doi:10.1016/j.giq.2016.09.002

Ginnis, S., Evans, H., Boal, N., Davies, E., Palmqvist Aslaksen, A., Ipsos MORI Social Research Institute, . . . Sciencewise. (2016). *Public dialogue on the ethics of data science in government*. Retrieved from London:

Graham, T. (2018). Data, nudges and design are changing the way we reach scale. *Scaling Social Impact.* Retrieved from https://apolitical.co/solution_article/data-nudges-and-design-are-changing-the-way-we-reach-scale/

Groves, C. (2011). Public engagement and nanotechnology in the UK: restoring trust or building robustness? *Science and Public Policy, 38*(10), 783-793. doi:10.3152/030234211x13070021633440

Gruzd, A., Jacobson, J., Kumar, P., & Mai, P. (2017). *Social Media Data Stewardship: The Ethics of Social Media Data Use for Research*. Paper presented at the Data Power 2017, Ottawa, ON.

Hancock, M. (2016). *Data science ethical framework launch: Matt Hancock speech*. London: Cabinet Office Retrieved from https://www.gov.uk/government/speeches/data-science-ethical-framework-launch-matt-hancock-speech.

Haste, H., & Hogan, A. (2006). Beyond conventional civic participation, beyond the moral-political divide: young people and contemporary debates about citizenship. *Journal of Moral Education, 35*(4), 473-493. doi:10.1080/03057240601012238

Ipsos MORI. (2016). *The one-way mirror: Public attitudes to commercial access to health data*. Retrieved from https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Jarrar, Y. (2017). What is the role of government in the digital age? *World Economic Forum*. Retrieved from https://www.weforum.org/agenda/2017/02/role-of-government-digital-age-data/

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional, 15*(6), 43-48.

Laurent, B. (2007). Diverging convergences - Competing meanings of nanotechnology and converging technologies in a local context. *Innovation-the European Journal of Social Science Research, 20*(4), 343-357. doi:10.1080/13511610701760804

Lauriault, T. P., & Mooney, P. (2014). *Crowdsourcing: A geographic approach to public engagement*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2518233

Mackinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

Malomo, F., & Sena, V. (2017). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. *Policy & Internet, 9*(1), 7-27.

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business: Report on a research agenda setting workshop*. Retrieved from Washington, D.C.:

McNeely, C. L., & Hamh, J. (2014). The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research, 31*(4), 304-310.

Michael, M., & Lupton, D. (2016). Toward a manifesto for the 'public understanding of big data'. *Public Underst Sci, 25*(1), 104-116. doi:10.1177/0963662515609005

Michaelson, J., Mahony, S., & Schifferes, J. (2012). *Measuring Well-being: A guide for practitioners*. Retrieved from London, UK: https://b.3cdn.net/nefoundation/8d92cf44e70b3d16e6_rgm6bpd3i.pdf

Milakovich, M. E. (2013). *Anticipatory government: Integrating big data for smaller government*. Paper presented at the Internet, Politics, Policy 2012, Oxford.

Park, A., Curtice, J., Thomson, K., Bromley, C., & Phillips, M. (2005). *British Social Attitudes: The 21st report*. Retrieved from London, UK:

Piscopo, A., Siebes, R., & Hardman, L. (2017). Predicting Sense of Community and Participation by Applying Machine Learning to Open Government Data. *Policy and Internet, 9*(1), 55-75. doi:10.1002/poi3.145

Poppy, G., & Food Standards Agency. (2017). *Chief Scientific Adviser's Science Report*. Retrieved from London, UK: https://www.food.gov.uk/sites/default/files/media/document/chiefscientificadviserssciencereport%20%281%29.pdf

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*. doi:10.1007/s12553-017-0179-1

Prolific. (2018). Prolific: Bringing people together to power the world's research. Retrieved from https://prolific.ac/

Qualtrics. (2018). Qualtrics. Provo, USA: Qualtrics. Retrieved from www.qualtrics.com

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. *Science, Technology & Human Values, 30*(2), 251-290. doi:10.1177/0162243904271724

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research, 31*(4), 343-348.

Sciencewise. (2012). Public dialogue on data openness, data re-use and data management: Sciencewise co-funded project 2012. Retrieved from http://www.sciencewise-erc.org.uk/cms/public-dialogue-on-data-openness-data-re-use-and-data-management/

Sciencewise. (2014). Big Data: Public views on the collection, sharing and use of personal data by government and companies.

Selin, C., Rawlings, K. C., de Ridder-Vignone, K., Sadowski, J., Altamirano Allende, C., Gano, G., . . . Guston, D. H. (2017). Experiments in engagement: Designing public engagement with science and technology for capacity building. *Public Underst Sci, 26*(6), 634-649. doi:10.1177/0963662515620970

Shelley-Egan, C., & Davies, S. R. (2013). Nano-industry operationalizations of "responsibility": Charting diversity in the enactment of responsibility. *Review of Policy Research, 30*(5), 588-604. doi:10.1111/ropr.12032

Stilgoe, J., Irwin, A., & Jones, K. (2006). *The received wisdom: Opening up expert advice*. Retrieved from London, UK: https://www.demos.co.uk/files/receivedwisdom.pdf

Stilgoe, J., Lock, S. J., & Wilsdon, J. (2014). Why should we promote public engagement with science? *Public Understanding of Science, 23*(1), 4-15.

Stirling, A. (2004). Opening up or closing down? Analysis, participation and power in the social appraisal of technology. In M. Leach, I. Scoones, & B. Wynne (Eds.), *Science and citizens: Globalization and the challenge of engagement* (pp. 218-231). London: Zed Books.

Stough, R., & McBride, D. (2014). Big data and U.S. public policy. *Review of Policy Research, 31*(4), 339-342.

Taylor, A. S., Lindley, S., Regan, T., Sweeney, D., Vlachokyriakos, V., Grainger, L., & Lingel, J. (2015). *Data-in-place: Thinking through the relations between data and community*. Paper presented at the HCI for Civic Engagement, Seoul, SK.

The European Social Survey. (2017). European Social Survey - Politics. Retrieved from http://www.europeansocialsurvey.org/data/themes.html?t=politics

The Royal Society, British Academy, Fourniol, F., & McLaughlin, F. (2017). *Data governance: public engagement review*. Retrieved from London, UK: https://royalsociety.org/~/media/policy/projects/data-governance/data-governance-public-engagement-review.pdf

van Deursen, A. J. A. M., Helsper, E. J., & Eynon, R. (2014). *Measuring digital skills. From digital skills to tangible outcomes project report.* Retrieved from www.oii.ox.ac.uk/research/projects/?id=112

Van Dyne, L., & Pierce, J. L. (2004). Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and

organizational citizenship behavior. *Journal of Organizational Behavior, 25*(4), 439-459.

Willis, R., & Wilsdon, J. (n.d.). *From bio to nano and beyond: A progressive agenda for technology, risk and the environment*. Retrieved from London, UK: http://www.green-alliance.org.uk/page_234.php

Wilsdon, J., Wynne, B., & Stilgoe, J. (2005). *The public value of science: Or how to ensure that science really matters*. Retrieved from London, UK.

Wynne, B. (2006). Public engagement as a means of restoring public trust in science--hitting the notes, but missing the music? *Community Genet, 9*(3), 211-220. doi:10.1159/000092659

CHAPTER 9

Additional Research

9.1 INTRODUCTION

This chapter presents a brief summary of results from external research projects that the author worked on in relation to government use of data during her doctoral studies. While all of these projects involved significant time and work by the author, they were collaborative projects not funded by the doctoral research itself. Thus, rather than including these results as individual chapters they are briefly summarised here. The two published papers related to this work are presented in full in Appendix C: Additional Published Works. These papers are important in understanding the results of Objective 2, as they include data projects conducted with local government. They are drawn upon in the discussion chapter in order to further understand the nature of government data use to promote public engagement practice.

9.2 FINDINGS FROM EXTERNAL RESEARCH

Table 9.1 presents results from external research as well as a brief summary of the projects and resulting papers. The key lessons presented below demonstrate that data practice in government is often politically contingent. In particular these political processes subject non-governmental actors, e.g. community organisers, commissioned service providers, and other non-governmental publics to the demands for the quantification of political decision-making. This was primarily demonstrated in the social prescribing project. In this project, data analysts from the University of Bath collaborated with local community organisations to utilise secondary data to help digitise wellbeing services in the local area. The publication (Rempel, Wilson, Durrant, & Barnett, 2017) that resulted from this work explores how social prescribing services were measured and subsequently how those measurements were used to assess whether the services were effective. This paper, and further reflection on the project as a whole, suggests that statistical analysis regardless of whether it is appropriate or cumbersome for the organisation, is seen as a vital tool in justifying commissioning decisions. This was true even in cases where the analysis was clearly inappropriate or impossible to conduct with

statistical validity. This does not suggest that service providers are manipulating the results or purposefully conducting poor statistical analysis but that commissioned services are strongly subject to the quantification of government processes regardless of their size or capacity. Oftentimes, non-statistically literate populations who provide services are required to 'speak the language' of statistics in order to justify or gain their continued funding.

Table 9.1: External research projects on the topic of government data use

Project Name	**Project Summary**	**Key Lessons Relevant to Thesis**
Collaborating to Deliver Social Prescribing in Bath and North East Somerset	In this project, local community third sector organisations, public officials, technical developers, and the research team, including myself, collaborated to explore the use of secondary data and digital technologies to develop wellbeing services in the Bath and North East Somerset area. I worked as a Research Assistant on the project, conducting and designing the research as well as writing up the results for publication. A paper resulting from this work was published in the journal *BMJ Open* and is attached to this thesis in Appendix C: Additional Published Works.	In this project, very limited data was available to the researchers and subsequently to local government to understand policy issues. Furthermore, data collection mechanisms, e.g. the choices of measurement instrument seen in the paper attached, was often poorly reflected upon. Community organisations needed to 'speak the language' of data in order to justify their commissioned status thus used a vast range of kinds of data to show evidence of good practice. These multiple and varied statistical analyses were less about having 'good data analysis' and more about using data to justify their position to local government in cases of funding scarcity.

Table 9.1 cont'd

Local government data integration project*	In this project secondary administrative data was used to explore and analyse various policy problems identified by policy officials in a local authority. This included a focus on how to leverage administrative data to inform service delivery in the context of education. I worked as a Research Associate on the project, collaborating to design the analysis. I conducted the statistical analysis on two out of four of the case studies and helped write up the results. The paper on this work has been submitted to a special issue in the journal *Politics and Governance* and is attached to this thesis in Appendix C: Additional Published Works.	Government data use, rather than just technical practice, is subject to political motivations. These motivations were related to three key domains: relevance, resistance and relationships of politics and data. In these processes, data practices were often shaped, shut down, or opened up by the policy questions at hand.

*The name of this project has been anonymised for confidentiality purposes.

The local government data integration project presents findings from a series of case studies on university/local government data collaborations that the university was a partner in. Similar to social prescribing, data projects were often driven towards assumed conclusions about what the appropriate policy decision would be. Key results from this project focus on the finding that data projects were not unproblematic technical processes, rather data was subject to various phases of establishing whether data was relevant, contesting data use, and contingent on trust. In conclusion, these studies found that local government data projects were

embedded in and part of complex political decision-making processes and more than data influencing political decision-making, politics influenced the kinds of data that were deemed relevant and appropriate to analyse.

9.3 Conclusion

This short chapter presents an introduction to relevant work conducted during the author's doctorate degree that was not part of the main programme of research. Both studies presented were part of a collaborative effort at the University of Bath's Institute for Policy Research to engage with policy-makers at the local level in data science. They were also opportunities to reflect, from an ethnographic perspective, on the practice of data science in policy-making. In conclusion, the projects demonstrate the strong preference towards quantitative decision-making in government practice even when the resources and skills for quantitative analysis are burdensome or unavailable. As well they discuss and reflect on the political nature of choices made in how to collect data, how to use data, and ultimately whether or not data analysis is politically relevant. They demonstrate the unique role of 'big data analysis' and 'data science' as a political tool.

CHAPTER 10

Discussion

10.1 INTRODUCTION

This chapter presents a discussion of the overall results from this thesis. First, a synthesis and summary of the results in relation to the aim and objectives of the thesis is presented. Second, the results are discussed in the context of the theoretical approach described in the methodology. Third, strengths and limitations are considered. And fourth, future directions of research are discussed.

10.2 A SYNTHESIS AND SUMMARY OF RESULTS

The aim of this thesis is to examine how publics and policy-makers can engage with emerging data technologies in the context of local and national government. This aim was specified through three objectives. First to summarise and evaluate the current research and state of government public engagement with data technologies. Second, to scrutinise the current use of data technologies in government through qualitative semi-structured interviews and other qualitative methods. And third, to examine the potential for new forms of public data engagement in government through qualitative workshops and analysis of public views on government and data technologies. The key findings from these three objectives are presented in Table 10.1 and discussed individually below. While each chapter presents a discussion of the results in relevance to the literature pertinent to that study, the purpose of this chapter is to examine how individual results can be discussed as a whole in relation to the aim and objectives set out. In doing so this thesis offers insight into both critical and traditional spaces for public engagement in how governments do or do not use data technologies.

10.2.1 Objective One

To address objective one, this thesis sought to explore the history and current practice of public engagement with government data use in the UK. Two key findings emerge from the studies in Chapters 3 and 4. First, that current public engagement practice is often limited in topic and form around assumed public concerns and second that public engagements with government data use should

better consider common themes from engagements in other science and technology areas.

Table 10.1: Main findings from thesis

Objective 1: To summarise and evaluate the current research and state of government public engagement with data technologies.	
Key Findings	(1) Current public engagement practice on government data use is often limited in topic and form around assumed public concerns like privacy and consent.
	(2) Public engagements with government data use should better consider common lessons from engagements in other science and technology areas, particularly avoiding public engagements that deliberately seek to create public trust.
Objective 2: To scrutinise the current use of data technologies in government through qualitative semi-structured interviews and other qualitative methods.	
Key Findings	(1) Data use in government is subject to political motivation particularly in the use of data to construct, test, and utilise narratives in policy decision-making.
	(2) Local governments, in particular, face stark challenges in adapting to civic datafication due to limited resources.
Objective 3: To examine the potential for new forms of public data engagement in government through qualitative workshops and analysis of public views on government and data technologies.	
Key Findings	(1) Publics express strong desires for engagement in multiple avenues and forms of inquiry.
	(2) Publics conceptualise of good engagement as flexible in topic and form, empowering for publics, and with meaningful potential to influence policy change.

This first finding is demonstrated in Chapter 3's discussion and summary of past government public engagements, where public engagement events were often about trying to establish the public view on data rather than holistic collaboration or decentred governance (see: Cameron, Pope, Clemence, and Ipsos MORI Social Research Institute (2014) and Ginnis et al. (2016)). Policy-makers and practitioners often framed the discussion in terms of assumed public concerns around concepts and practices related to privacy and consent. Engagement was then a process of addressing these assumed concerns in order to allow data science practice to expand and flourish. As discussed in Chapter 4 in the ethnographic exploration of one particular public engagement exercise, policy-makers were explicitly focussed on constructing the guarantees around data use needed by publics in order to provide a safe space for data scientists. Public engagement exercises also functioned as safeguarding exercises for government data practitioners as they aimed to develop a social licence for data science to proceed unimpeded by public concern or protest (Carter, Laurie, & Dixon-Woods, 2015).

A potential explanation for the nature of this limited focus is the connection highlighted in Chapter 3, that the aim of engagement and the structure are closely connected. The data engagement activities reviewed in this chapter were often one-off, small group discussions and this is likely because of, and subsequently contributes to, the aim of receiving comments on a narrow set of issues related to data technologies. These were not exercises that encouraged public opinion on live policy or data problems. This thesis contends that these kinds of one-off, limited scope events are invariably open to the critique that they are tokenistic. Public engagement exercises around data rarely demonstrate mechanisms for establishing change in data technology or in policy development. While this is a confirmatory finding for data engagement practice, in that other forms of novel governance practice like the formation of mini-publics to assess and recommend policy options are commonly critiqued for tokenism (Setälä, 2017), it further substantiates that public engagement is not perceived of within government as an exercise that allows differing and critical public voices to influence data practice.

This thesis finds that public engagement exercises are limited in their scope to change, or indeed simply comment on, how governments employ data technologies.

While it is evidently important to consider topics like privacy and consent in data usage, public engagement on data technologies, as a practice employed by policy-makers, is thus far focussed on identifying and calming concerns rather than opening debate. Public engagement topics are subsequently narrow and divorced from data practice. This finding is echoed in work by Felt and Fochler (2010) who similarly describe public engagements with science and technology as exercises in "surveying and assessing potential critical voices" (p 228). These are not exercises in shifting power away from central structures. Rather, public engagements with government data use often align with Fiorino's (1990) description of instrumental purposes where publics are included as a means to an end, e.g. to make a space safe for data scientists, and not to influence practice.

There are two foci within the scholarship and agenda of public engagement that diverge on whether engagement on any technology topic is about consulting and calming or collaborating and governing. The first focus is that engagement practice does not need to focus on establishing influence, and as seen in the previous paragraph it often does not, but rather it can focus on capacity building of publics, e.g. as knowledge and skill building exercises (Selin et al., 2017). The second focus is that public engagement is an activity of governance, in that publics must have an established influence over the development, regulation, or assessment of new technologies (Groves, 2011). The focus of public engagement as capacity building can be seen in things like civic hacking and citizen science (Lauriault & Mooney, 2014; Schrock, 2016). Publics operate as data collectors, data analysers, and other forms of scientific actors on a variety of policy and scientific areas. The value of taking part in public engagement is learning how to perform in these capacities (Dickinson et al., 2012; Riesch, Potter, & Davies, 2013). Evidently then the scope for what 'valuable' public engagement looks like is expanded

beyond the discussion-based and deliberative exercises seen in Chapter 3's Dialogue on Data Science Ethics. Also expanding the form public engagement takes is the growing literature around *using* data technologies to engage publics, for example social media, that demonstrates a renewed focus on moving public engagement beyond the one-off, small group discussion format (Lee & Kwak, 2012). This thesis, from a feminist perspective of establishing voice and recognising the limitations of power that decentred governance seeks to address, contends that while capacity building is a laudable and often useful exercise it is a limited venture that does not address larger societal issues around the power embedded in the use of data in government. In fact, the aims of public engagement and the aims of citizen science are often divergent (Riesch et al., 2013). Thus, forms of distributed work seen in citizen science and civic hacking should not be equated with substantive public engagement on how governments use data. Furthermore, while there are novel forms of public engagement being developed, digital or otherwise, these engagements are rarely if ever on the topics of data use itself (Lauriault & Mooney, 2014). Thus, this thesis finds that while government data use is accepted as topic of debate within the UK, these debates are still limited to strict topics and forms that do not challenge existing power structures.

The second key finding related to Objective One is that public engagements with government data use should better consider common themes from engagements in other science and technology areas. While it is encouraging, from a feminist perspective, that government are interested in finding different public voices on data science, it is clear from the literature review and the ethnographic study that these engagements are still small and constructed events that are facing similar challenges experienced by other public-facing technologies. For example, assumptions around information equating public acceptance is evidenced in the Care.Data and Google DeepMind cases discussed in Chapter 3 (Carter et al., 2015; Powles & Hodson, 2017). In these cases, government data practice was critiqued for not effectively considering publics in their sharing and analysis of personal

data, and relying on the argument that by informing publics through leaflets and consent forms about their practices that publics would be comfortable in the use of personal data by third parties. In early science-citizen interaction from the 1980s, e.g. nuclear power, there was similarly a phase of public engagement where government and industry assumed that more information would lead to better acceptance of science and technology, but this was evidently not the case. In fact, this deficit model, i.e. that publics hold a deficit of knowledge, created further controversy as it suggested publics were unable to hold valid concerns related to science and technology. Rather than learning from this history, this thesis finds similar debates arising on the topics of public engagement on government data use without critical reflection on how to approach public engagement form a more nuanced perspective.

There is, in general, a limited focus on data-related public engagement from both government and public engagement scholars alike. While this is partly due to data technologies being relatively novel sociotechnical objects, Stilgoe et al. (2014) critique that public engagement literature has thus far been preoccupied with engagement in the biological and life sciences. Similarly, Michael and Lupton (2016) in the journal *Public Understanding of Science* call for renewed focus on public conceptions of big data technologies. Traditional practitioners of public engagement in the UK like Sciencewise and Ipsos MORI have indeed run several public engagements on data-related concepts in recent years (Cameron et al., 2014; S. Davidson et al., 2013; Ginnis et al., 2016; Ipsos MORI, 2006, 2016), however these exercises do not evidence that government have reflected on how past mistakes could inform future practice. As described in the first finding, government are still focussed on finding safe spaces for themselves rather than enhanced spaces for publics. Governments in the UK are not reflecting on lessons from past public engagement practice and public engagement practitioners are also not necessarily focussing on government data practice.

Further in line with this lack of reflection on past practice is the dialogue from a recent conference on the topic of critical data studies (Data Justice Lab, 2018). One of the participants tweeted that data practice faces a 'crisis of trust' (Hintz, 2018). This theme of the discussion further evidences that the narrative of the relationship between practitioners and publics in the context of data is mirroring issues of the past. As discussed in Chapter 3, the UK House of Lords (2000) made this same assertion nearly two decades ago. Specifically, that due to concerns around GMO and BSE technologies the UK faced a 'crisis of trust' in science. This led to a decade of public engagement that focussed on building trust, but as critiqued by Stilgoe, Irwin, and Jones (2006, p. 20) "the deliberate attempt to manufacture trust can look deeply untrustworthy". Public engagement on government data use will face this same challenge if practitioners do not reflect on avoiding limited engagements that are seeking to be merely trust-building exercises. Subsequently, this thesis argues for a better integration of public engagement scholarship, critical data studies, and government practice. In particular, for public engagement with government data use to be seen as an ongoing practice that is embedded in data practice itself rather than one-off exercises that focus on establishing the 'public view', independently from data practice and with no clear line of influence.

In sum, the findings from the first objective of this thesis highlight that current data engagement practice is limited in scope both in terms of topic, i.e. what is up for debate, and influence on practice, i.e. the power publics have to decide how their data and other data is used by government. Further, these data engagements are mirroring the mistakes and challenges faced by past technologies without critical reflection from the academic community nor from government. While there is a growing emphasis on allowing space for public engagement in government data use, these engagements do not reflect a substantive commitment to including the public in any aspect of data technology development or governance.

10.2.2 Objective Two

In Objective Two, this thesis sought to scrutinise the current use of data technologies in government. In doing so, discussion of what this means for public engagement practice is also presented. Two key findings emerge from this objective in Chapters 5 and 6, as well as the external research presented in Chapter 9. First, that data use in government is subject to political motivation particularly in the use of data to construct, test, and utilise narratives in policy decision-making. Second, that local governments, in particular, face stark challenges in adapting to civic datafication due to limited resources.

The first finding that government data use is subject to varying forms of political motivation is a common thread throughout the thesis. It was triangulated through different methods and studies including the story completion exercises in Chapter 5 and the semi-structured interviews in Chapter 6. A major contributing theme from the story completion exercises in Chapter 5 found community organisers needing to align individual and organisational aims for data sharing to occur. This aligning of aims suggests that data sharing is contingent upon political acquiescence. It is not simply following set data ethics procedures but also subject to nuanced and complex projections of whether data sharing would benefit local government. Also contributing to the finding that data use is politically contingent was the theme in Chapter 6's interview analysis around the use of data in local government to construct narratives. Data analysis was used to develop, confirm, or contest policy narratives. In many ways, data analysis was impossible to separate from the mechanisms of shutting down or opening up narratives. Thus, data analysis is not simply objective evidence that exists independent of the policy context but is intertwined with the actions and aims of policy-making. Also relevant is the discussion around the political motivations of data analysis seen in the external research papers in Chapter 9. This chapter discussed various case studies relating to data projects with local government. In these projects, the use of data in policy was not simply a process of incorporating evidence in decision-

making but rather data use followed complex pathways of determining the relevance of data and resisting data use where it did not support the narrative. All of these themes demonstrate that data projects are not only technical processes in government but subject to the nature of the policy problem at hand. For example, as described in Chapter 9's case studies of local government data integration projects, community organisers did not necessarily approach data analysis as an objective exercise. Rather they saw data analytics as an organisational risk that funders and commissioners would use to justify the de-funding of community programmes in a resource scarce environment. Data analytics, data science, and other aspects of civic datafication are thus highly politically contingent.

The design of data projects and their ultimate success rests on what is politically feasible and desirable. The notion that policy-making is a process of iterative narrative testing and narrative advancing is not a new one, see the discussion in Chapter 2 around Deborah Stone's (1989) work dating back several decades. However, the finding that data is not simply a technical solution to policy-making is not well recognised in policy-making processes nor is it often discussed in academic literature. While there are some authors who discuss data as influencing the policy-making process (see: Janssen and Helbig (2016); Kettl (2016); Milakovich (2013)), this thesis finds that policy narratives influence data processes. Thus, data is not only a tool to promote or refute narratives in policy-making but is also subject to the contingencies of political motivation and desirability. While it is not surprising that data is subject to politics in this manner, this thesis adds nuance to the ways in which data science at every level from creation to analysis to policy-making is curated by the priorities and contingencies of government. This thesis adds to a growing literature that contends that data science procedures as deeply subjective processes. The results of data analysis are dependent on subjective and political decisions around appropriate and desirable measurement, analysis and interpretation (Johnson, 2014; Milakovich, 2013). Data science and subsequently civic datafication are unique in that they are not simply 'evidence' for political decision-making that can be taken up or ignored dependent

on political aims but are political activities in and of themselves. The act of data science becomes the act of politics.

The second finding highlights the practical realities of local government adopting data technologies. In Chapter 6, interview participants highlighted several limitations related to finite resources, costs, time, and skills to use data effectively, as well as the siloed nature of local government administration. Furthermore, Chapter 5's story completion exercise demonstrated that even when imagining relatively simple data sharing projects between service providers and government commissioners, policy officials struggled to imagine a positive outcome, in other words one where effective data sharing took place. And again in the external research project on social prescribing described in Chapter 9, in practice it was very challenging to find secondary data that would be of use to local government. This limited capacity in terms of funds and resources of local government, and governments in general, is discussed in the literature by authors like Kitchin (2013) and Malomo and Sena (2016). Few, if no, articles examined in this thesis suggest that government is fully technically capable of adapting to complex and large-scale data processes. In fact, Orr and Vince (2009) demonstrate that UK local government, in particular, has a historical tradition of crisis in which local governments are "in states of chronic stress, and existing in a hostile context, especially in one of increasing central control" (p. 668). This narrative of limitation and lack of resource is common. However, it is important to note that this narrative, although discussed theoretically in the literature around things like smart cities (Kitchin, 2013), is not often empirically explored in the data context as is done in this thesis.

The finding that scarce resources are fundamentally tied to the ability for local government to adapt to civic datafication is particularly valuable in demonstrating that the tradition of crisis has in fact both driven the uptake and lack of uptake of data technologies. In other words, the tradition of crisis has driven both the perspective that data technologies will help cut costs and save on resources, and also the lack of uptake of data technologies, i.e. governments do not

in fact have the resources to employ data scientists. These coexisting themes are a novel finding of this thesis. They are able to co-exist because of the disconnect between the perceived positive rhetoric around civic datafication and the practical reality of limited funding and resource in government. Central government continues to push the narrative that data is the salvation of social and government problems (Jarrar, 2017; Manzoni, 2017). If this narrative is to succeed, there must be a commitment to supplying the resources, both human and financial, to ensure that data projects can happen.

This thesis now reflects on the relevancy of the previous two findings to understand the capacity for governments of any size or structure to engage publics. Considering the lack of resource for data projects in local government, it calls in to question whether public engagement is indeed feasible. If governments do not have the capacity to run data projects or commit to a model of 'digital by default' then public engagement on these processes will need to recognise this limitation in order to be effective (Andrews et al., 2016). This thesis contends that public engagement must therefore focus both on means and ends. As local and national government are challenged to adopt data technologies under financial and resource constraints, there is an opportunity for publics to be involved in deciding the means of data usage. In other words to be engaged, on the kind of data that is used and developed. These are questions of what data technologies would be effective, what kind of data is collected, and how data is ultimately used. Engaging on means can also help address the inherent power relations that develop from governments solely deciding what is relevant to measure, how to measure that thing, and then how to use the results from that measurement. Longo, Kuras, Smith, Hondula, and Johnston (2017) discuss a key example of these imbalanced power relations in how digitisation of government creates a class of individuals who are digitally invisible, e.g. homeless populations. These individuals are not using the technologies that render them visible to government in the modern age, e.g. smartphones. Public involvement with these individuals is imperative in order to both empower their voice and to help government

anticipate what kinds of measurements and subsequent service provision would be relevant to their needs. These processes of engaging publics can help government focus resources on what is most important to track and measure for their local area. Engaging publics on means can be mutually beneficial.

As well as means, this thesis also finds relevance for public engagement to consider the ends of data analysis. Data projects do not exist in a vacuum. As is argued here, data projects intend to do political things. Data projects impact and are driven by political motivations. Thus publics can be engaged on these political ends, i.e. what data projects intend to do. For example, in the case of the social prescribing project, measurement was a political act of deciding what was relevant and appropriate to measure. Measurements that are perceived to be politically preferred, e.g. funding and cost-benefit analysis, do not necessarily reflect the values of social prescribing which is focussed on the social aspects of wellbeing. While there is a wealth of literature to draw on related to distributed democratic processes (Setälä, 2017), this thesis highlights that public engagement practitioners and policy officials should not lose sight of the ultimate goal of data projects, i.e. to make political decisions. In Chapter 6's interview analysis, this thesis found that data, as a term, is often drawn upon in decision-making processes for its perceived objectivity. This objectivity was thought to calm the emotional side of policy-making. The notion that data can be a tool in calming opposition to policy decision-making can only be problematised through a broader inclusion of diverse perspectives and views on policy decision-making. Public engagement on data can further the acknowledgement that data can only tell part of a story in a policy decision. In the case of Chapter 9's social prescribing example, this would involve engagement exercises both on how to measure social prescribing but also on what good social prescribing practice looks like and how it should be operationalised in community health practice. Public engagement must engage publics both on what and how to measure, as well as on the ultimate use of those measurements to make decisions related to political aims.

The findings around Objective Two suggest that there are both policy and data questions to be answered in the context of civic datafication. Civic datafication is not just about the technical adaptation of government to the future but how data influences the relationship between publics and government. Public engagement on government data technologies must reflect all aspects of the civic datafication model, i.e. both means and ends. From the perspective of decentred governance (Griggs et al., 2014), public engagement must acknowledge how data technologies impact and have the potential to draw out different pluralistic public voices but also how government data use shifts the power of decision-making away from publics. Public engagement in the context of civic datafication should see engagement as a process of problematising both the policy decisions that are made using data and the mechanisms through which data is designed, collected, and used.

10.2.3 Objective Three

In addressing Objective Three, this thesis seeks to understand the nuances of how publics wish to be engaged, through both qualitative and quantitative analysis. Specifically, to examine the potential for new forms of public data engagement in government through qualitative workshops and analysis of public views on government and data technologies. Two key findings are discussed below. First that publics have strong desires for engagement in multiple avenues and forms of inquiry. Second, that publics conceptualise of good engagement as flexible in topic and form, empowering for publics, and with meaningful potential to influence policy change.

In the survey study in Chapter 8, participants expressed a strong desire to be engaged in all forms of data engagement presented to them including communication, consultation, and collaboration. Within the UK, this desire to be engaged is evidenced through public events reported in the media, see Care.Data (Carter et al., 2015). It is easy to assume from these kinds of events, that publics main desire for engagement exists around the ability to keep their data private.

However, the results found in Chapter 8's mediation analysis suggest that although engagement preferences were statistically related to both feelings of ownership of data and type of data, ownership was not a mediating factor between type of data and engagement preferences. In other words, although participants wanted to be engaged around more sensitive, passively created kinds of data like health information these desires did not relate to wanting to own, and subsequently keep private, that data. In general, participants expressed the desire to be engaged on all types of data and through a multitude of mechanisms. This included, although to a slightly lesser degree, helping government use data better. Subsequently, the 1100 UK residents in Chapter 8's study demonstrate that publics may be open to seeing data engagement as an activity that could benefit their local community, government, and themselves.

While, this thesis finds strong evidence of public desires for further representation within government data practice, it is evidently important to reflect on the feasibility of this kind of long-term and ongoing engagement. While the history of public engagement with science and technology discusses the key role publics should play in debating the relationship between technology and society (Burgess, 2014), literature located in the political sciences often conflicts on how effective decentred governance can be, and indeed whether a stronger representative democracy should be the prevailing model of government-public interaction (Achen & Bartels, 2016; Farrell, 2014; Galston, 2010). For example, Galston (2010) contends for a more nuanced and realist view of political theory in which idealised views of the democratic process and subsequently representivity are rejected. Further complicating the feasibility of enhanced public engagement is the prevailing perception outlined in Chapter 7's focus group analysis that publics and government imagine the 'public' to be disinterested and unintelligent in relation to data practices. The finding here that publics have a strong interest in being engaged contradicts this perception. This suggests one mechanism through which to increase the feasibility of public engagement, is to develop models for reimagining the ways in which public engagement is perceived within

government. Public engagement needs to be presented, and encouraged, to government and publics as a collaborative exercise that reaps mutual benefit.

The second finding highlights the nuances of what publics in Chapter 7's workshop study perceive good engagement to be. Specifically that public engagement should come in many forms, focus on means and ends, empower publics and be influential in policy-decision making. The themes of having multiple in-roads for participation and the need for engagement exercises that impact decision-making are often discussed within both the public engagement literature as well as theories of decentred governance (Burri & Bellucci, 2008; Setälä, 2017). It is also important to note that participants in Chapter 7's focus group study did not necessarily want technical solutions to the political issues of data use and data-based decision-making. In fact participants in Chapter 7 were critical of online platforms for local participation like websites or apps. Participant's desires for engagement focussed less on the practical mechanisms for participation and more on the nature of the relationship between publics and their government. It was challenging for participants to imagine a different way of including publics than the traditional means of one-off consultation exercises. All three public groups, government, non-technical, and technical alike, were sceptical of the success of decentring the governance of data technologies. This scepticism was based on the perception that there was a combative and disengaged relationship between government and the citizenry. Evidently the first step to decentring data governance, is to address these kinds of social perceptions. In order to address these perceptions, this thesis takes note of the point made by the participants here and in the public engagement literature that engagement must include multiple different methods for public participation. These different methods must take note of the different levels of data literacy in the citizenry as well as the administrative and long-term nature of government data projects. In sum, while publics are strongly interested in being engaged around data technologies, publics and government alike imagine the opposite. Public engagement on government data use must adapt to these perceptions by offering

multiple methods for public participation, an engagement model that is based on mutual benefit, and genuine potential for engagement that influences data-based policy decisions.

10.2.4 Final key contributions to knowledge

Finally, discussion on the key contributions to knowledge is presented through discussion on the critical and traditional spaces for public engagement in government data practice. For the purposes of this thesis, the term public spaces is used to describe the imagined and actual allowances that publics have in government data science processes. Traditional spaces are those that are established and to some degree uncontested within government whereas critical spaces are those which offer new opportunities for publics to be involved in government data practice and more broadly in civic datafication.

Traditional spaces found in this research include established mechanisms of privacy and consent, public engagement through small group discussion, and consideration of the 'public good'. Consent seeking and privacy regulations are the most basic established allowance for publics in government data practice. Consent for data usage, if considered as a form of public engagement, would exhibit a limited kind of consultation (Rowe & Frewer, 2005). In other words, while publics are being asked if their data can be used, the terms of that use is neither open nor debatable. Self-evidently in the case of government service provision it is also a non-optional consent in that one must provide consent in order to engage with the government service in question. While there are more dynamic and flexible forms of consent being developed (Teare, Morrison, Whitley, & Kaye, 2015), nevertheless consent practice in government is currently an instrumental activity that does not permit publics to have influence on the debate over civic datafication (Fiorino, 1990).

A second traditional space is small group discussion public engagement, as presented in Chapters 3 and 4 government has begun to engage publics on topics like data science, machine learning, and personal data sharing. As critiqued in this

thesis, however, these exercises are often divorced from practice and indeed are generally about safeguarding government data practice. The third traditional public space exists in official guidelines put forward by central government suggesting data practitioners should keep the public in mind by ensuring they focus on data usages that are in the 'public good' (Ginnis et al., 2016; The Royal Society, British Academy, Fourniol, & McLaughlin, 2017). While these traditional spaces demonstrate that governments at the local and national level are open to considering publics in the governance and development of data technologies in policy-making, they are not areas in which publics have the capacity to influence larger debates because they separate the act of government data practice from publics. These traditional spaces can be easily critiqued for tokenism in which the public is included to assuage potential concern rather than to debate and advance how and why data is used in government (Fiorino, 1990; Selin et al., 2017). That is not to say that these spaces should be shut down, in fact this thesis contends that publics express a strong desire to be asked and told about the use of their personal data. However these current traditional spaces in which publics are asked or told about data usage is approached by government as a way to shut down public influence rather than open up the opportunity for practice to change. Consent does not provide an avenue through which individuals or publics can voice thoughts or concerns. Current public engagements are disconnected from actual government data practice. And, keeping the 'public good' in mind is impossible if publics are not permitted to debate what the public good means in practice. In order for these traditional spaces to problematise power relations between governments and publics in the use of data, they must be re-thought of as exercises in substantive public inclusion rather than government safeguarding.

Critical spaces are new opportunities for public engagement with government data practice. They are not necessarily better suited to stronger inclusion but as these are new spaces, they offer a better opportunity to be developed under a model of substantive public inclusion from the start. This

thesis finds critical spaces in debates on imaginations of the future and in the practice of government data use.

A first critical space is in the imaginations of civic datafication. Specifically, as found in Chapter 3 and further reflected on in the discussion around government data practice in Chapters 5 and 6, exercises in public imaginations of the future offer the opportunity for publics to discuss what data should be collected, how it is used, and what policy decisions are drawn from it. These should be long-term and established practices; not one-off events as is often the case in consent seeking and traditional public engagement. Rather than critiquing publics as unable to understand the complexities of data technologies as was the case in a public engagement exercise presented in Chapter 3, publics can be engaged around debates of values. Engaging publics on imaginations can help address the challenge of many publics not having the technical expertise to discuss data technologies. One method presented in this thesis, the story completion method, has the potential to be adapted to enable a process of asking about narratives and imaginations rather than practice. The goal of engagement should not be to create technical experts out of publics rather it should be to find mechanisms of engagement that do not classify non-technical publics as less important in the decisions made around government data practice. In this way public engagement can help address the ways in which digital technologies and data practices introduce problematic power relations between government and publics.

A second critical space is the practice of data science in government. Critical, inclusive, and substantive public engagement requires governments to include publics in what established data practices consist of. Publics need to be engaged early in how those systems are built, what those systems do and how that impacts day-to-day life. This poses a challenge to local and national government in the UK to be more transparent about what those kinds of data systems consist of as well as better tracking their own use of data. This kind of engagement should

again be more than one-off and short-term exercises. This thesis does not contend that government should attempt to establish the 'public view', as has been done thus far in traditional spaces, because that is not a question that has a simple, single, nor unchanging answer. Rather from a model of pluralism, governments should enable mechanisms, within current data processes, that allow publics influence into how their data is used and what sorts of things large scale data projects can do to address social problems. Data technologies have the potential to do great good, but these benefits will be under realised as long as government and publics are not equally powerful in ensuring these technologies are debated, regulated, and put to socially sanctioned purposes.

10.3 CRITICAL AND THEORETICAL REFLECTIONS

In Chapter 1 this thesis presented several underlying arguments as to the motivations behind this research. These arguments were underpinned by theoretical reflections of feminism and decentred governance as described in the Methodology chapter (Davis, 2008; Ferguson, 2017; Griggs et al., 2014). This section discusses whether the results presented above support or refute these arguments for enhanced inclusion of publics in government data use.

1. The first argument was that data use and data technologies are developing faster than our social and critical evaluation of them. This argument is supported by the results from Chapters 5 and 6 in which local governments struggled to negotiate what was and was not ethical in their own use of data. In fact, government officials discussed a strong hesitancy to share and collaborate with data even within their own institution due to a lack of confidence in what was and was not permissible. Similarly, data events like the Care.Data sharing plan described by Carter et al. (2015) and the Google DeepMind commercial-government data sharing exercise as described by Powles and Hodson (2017), demonstrate that social and critical evaluation has not kept up with the technologies themselves.

Social and critical evaluation of data technologies is still in its infancy, and publics are limited in both opportunity and scope to debate data technologies.

2. The second argument was that publics are often unaware both of the kinds of data analysis conducted on the data that they have (albeit often unknowingly) provided and of the impacts of that analysis by government, academia, and industry. In the workshops presented in Chapter 7 participants actually found it quite easy to point to the many ways in which data is collected on their day-to-day activities. However, non-technical and technical publics were critical of whether government effectively used data and indeed struggled to point to examples of government data use. Reflecting a lack of awareness by publics on government data science, Ginnis et al. (2016) found that 68% of 2003 adults in Great Britain had never heard of or heard of but knew nothing about the term 'data science'. This provides some support that publics are not aware of data practices. However, following large-scale media events like the Facebook Cambridge Analytica scandal (Adams, 2018) this thesis indeed finds that public awareness is quite strong in relation to the datafication of day-to-day life.
3. The third argument was that individuals and publics have a fundamental right to know how their data is used and how that impacts both themselves and their communities. This argument is rooted in the issue of power (Davis, 2008; Ferguson, 2017; Griggs et al., 2014), in that without knowledge publics cannot address how they want, or indeed may not want, practice to change. Chapters 7 and 8 demonstrate that publics are indeed interested in the governance of their information, however Chapters 3 and 4 demonstrate the public role in government data use as one in which deference is paid to the 'public good' but not public opinion. Publics

are not substantively engaged. Put simply in The Royal Society et al. (2017) review of public engagement for data governance, "[t]here is a need for more engagement and communication about data and its use" (p. 4). Thus while this argument finds strength in the interest of publics, as well as the complex and contested nature of data usage described throughout the thesis, the reality is that public engagement on data technologies requires further feminist scholarship that simultaneously promotes and studies this topic.

4. The fourth argument was that democratic governments, due to their mandate as both representing and serving publics, should be a leader in ethical and transparent data use that takes in to account public views and opinions. As discussed in Chapters 3, 5, and 6, governments face great challenges in deciding what is and is not ethical in their own use of data, and as shown in Chapters 7 and 8 publics are interested to varying degrees in having a say in that usage. Governments like those in the UK are responsible to these publics. From an idealistic view of democracy, in other words a kind of democracy that can be criticised for being impossible (Galston, 2010), publics would elect governments that best represented their ethical views and desires. Such a utopian government, evidently, does not exist. However, governments in the UK purport that they do want to be responsive to publics, what this thesis furthers is that government have a mandate to move beyond considering the 'public good' and instead to a model of inclusivity and the decentering of the governance of data usage.
5. Finally, the fifth argument was that data technologies, for example surveillance using big data, should be understood as social objects that must be put under public and academic scrutiny. This final argument is positioned as the rallying cry, as it were, for this body of work. This thesis introduces and promotes the concept of terms like

civic datafication, as well as promoting an agenda of public engagement, with the hope that they will encourage better consideration of the social aspects of data technologies. It would not be an overstatement to say that government in the UK is committed to creating a form of government that is based on algorithmic justification and guidance, in other words 'digital by default' (Cabinet Office & Government Digital Service, 2017). And indeed academia is interested in the use of data by both government and private organisations (Data Justice Lab, 2018), however, this area of research is still new and finding its focus. This thesis hopes to centre some of the future areas of research by focusing on the relationships between publics, government, and data technologies from a contextualist and feminist perspective. The three publics represented in this study, e.g. non-technical publics, government publics and technical publics, are unequally powerful in deciding the future of civic datafication within the UK. This thesis demonstrates this unequal power while also promoting methods and concepts that can help even the power relations between these different publics. Future research should reflect on and promote public engagement change through focus both on practice and theory, on change and imagination.

10.4 Overall strengths and limitations

This thesis' main strengths lie in its interdisciplinary and pragmatic approach to research. The findings related to each objective are examined through multiple studies and several different methodological approaches. As well due to the exploratory nature of the work, many of the results are novel, particularly the conclusions in relation to public preferences for engagement as well as the nuanced examination of how data technologies are used in government. However there are several limitations to recognise in this approach. The diverse and

distributed nature of the topics covered means that there is likely previous scholarship and current ongoing work that is not drawn upon or discussed. The main limitation to an interdisciplinary approach is that it risks addressing each of its topics to a shallow depth without any deep exploration of a single topic. This thesis has not drawn on the literature related to governance or feminism to the degree in which the author would wish. However, it was decided that this thesis would engage deeply with the topic of public engagement within the United Kingdom, which is itself interdisciplinary, and it is hoped that this level of depth is sufficient for meaningful and interesting scholarship. This thesis sees the benefit of interdisciplinary research being this bringing together of different concepts from different fields rather than finding novel results within a single discipline.

Finally, the qualitative nature of much of the thesis may lead some to question how generalizable and reliable the results are, in other words the quality of the research. This thesis did not approach qualitative analysis, specifically thematic analysis, as an activity of verifying results between independent coders. All of the analysis was conducted by the author, with a phase of confirmation and refining of themes with the supervisory team. As outlined by Vasileiou (2015) qualitative research quality guidelines vary greatly but common themes include transparency of methods, whether substantial evidence was provided for the conclusions drawn, and reflexivity on the author's positioning in the research. The author hopes that the themes and conclusions drawn are sufficiently supported by the quotes and statistical analysis presented. In addition as Braun and Clarke (2006) describe, themes are not meant to be incontestable, in fact there will likely be contradictions with the data. As well the author has been reflexive and reflective throughout the thesis, including the presentation of arguments and motives in Chapter 1, the reflexivity section in Chapter 4's ethnography and finally in a concluding reflexivity discussion following this chapter. As this research is exploratory, the author argues that instead of the analytic approach being a limitation it is a challenge for further research in the future to examine whether the themes found here are present elsewhere.

10.5 FUTURE DIRECTIONS OF RESEARCH

While, like any project, there are several lines of research that could build on this work, there are three areas in particular that would most benefit from further focus. First, establishing further cases addressing the same research questions. Second, building on the final study to test new and critical forms of public engagement in data-related policy-making. And third, opening the context of the research to examine private industry data usage.

The first area of further research is related to the contextualist nature of the thesis. As this research was approached as a knowledge-building exercises through case studies, the author encourages further research on these same topics. Further study of the nature of data use in governments as well as the contingencies of public engagement practices could demonstrate whether the themes from here are reflected in other governments and with other publics. Broadly conclusions like the recommendations made in Chapter 3 regarding suggestions for government public engagement exercises could be further developed and tested in similar and diverse contexts. Particularly, further exploration of public preferences for engagement in regard to feasibility would be useful. For example, studying the feasibility of local government creating public engagement resources as well as public perceptions of such resources. Rather than this thesis being viewed as the final word on the relationship between publics, policy-makers, and data technologies, this thesis is meant to open up debate on the novel concept of civic datafication. Many of the conclusions and themes discussed here were exploratory and thus there is now a programme of work needed to further examine and add to these results.

Second, future research should also test new and different models of public engagement. While this thesis focussed on establishing a base on which public engagement on data usage could build, further research should encourage novel design of public engagement mechanisms themselves. In particular, engagement exercises that take into account the themes and conclusions from the various

chapters is an important next area for research. These exercises should be dynamic, long-term, embedded in the data work itself, and responsive to publics' ongoing feedback. There are numerous challenges in incorporating long-term forms of public engagement in government thus this area, in particular, requires empirical and explorative case work. This thesis would suggest a first step being further development of online resources that are simple to use for both publics and local government to list both current projects related to data use as well as areas of interest regarding data to different publics.

Finally, while civic datafication focuses on the relationship between data technologies, government, and publics; the use of data technologies is not unique to government. Further research should examine mechanisms for distributed governance of data usage in private industry. Private companies often look to government to set the rules for acceptable and ethical technology usage (Walls, O'Riordan, Horlick-Jones, & Niewohner, 2005), thus exploration of private industry's willingness and capacity to engage publics should be a key area of research. While this thesis argues that governments exist in a precarious space in which they must regulate themselves in good data practice, government also must regulate private actors in the appropriate inclusion of publics in data management. Thus a natural next area of research is how government regulates and monitors the use of data by companies like Facebook, Google, and related big technology subsidiaries.

CHAPTER 11

Conclusion

11.1 PERSONAL REFLECTIONS AND FINAL THOUGHTS ON REFLEXIVITY

In this section I discuss various reflections and challenges faced during the research, in particular the challenges around inhabiting an interdisciplinary space in modern academia. As presented in the introduction and throughout the thesis, this research was approached from several different disciplines. I studied for my Masters and Undergraduate degrees in an interdisciplinary Health Sciences faculty in Canada. I came to Bath having studied, albeit with varying intensities, in the fields of health research, molecular and microbiology, health ethics, medical education, public policy, forensic psychology, chemistry, archaeology, writing, Indigenous studies, actuarial sciences, and epidemiology. This is not presented as evidence of a great depth of understanding in these studies, but rather as indicative of the sort of researcher I am trying to become. I have always sought to draw things together and learn from as many different kinds of people and thus disciplines as possible. I brought these sentiments to my doctoral studies; however I have personally found British academia to be an immensely different place than I had experienced in Canadian academia. Here, disciplines are closely guarded. Thus I found myself at the forefront of being an interdisciplinary PhD student, in fact a social health sciences student in a psychology department, having come from a place where that was not a challenging identity to hold.

This meant that I was faced with learning how to do interdisciplinary doctoral research without the support of a university that values interdisciplinarity and particularly not interdisciplinarity research with predominately qualitative methods. Practically this meant that I was learning how to do many new things related to qualitative research without the support of formal teaching arrangements. As well I had to navigate a vast literature on policy and public engagement in the UK context that in Canada I was wholly unfamiliar with. Personal challenges aside I also held strong views on advancing the cause of publics in government data usage, hence my findings and the areas in which I chose to research are inevitably going to reflect my perspective that indeed publics should be involved in civic datafication. I have tried to be as transparent as

possible in this by presenting these views and arguments in the Introduction chapter. From a critical perspective, I hope that my lack of familiarity with the UK system of academia and government, as well as my feminist interest in public engagement, led to me discovering different aspects of public engagement and policy-making than would be found by someone without my background.

Thinking more broadly, British academia and any singular university is far from being alone in their focus on pure disciplinary studies. The world of academic publishing is still a place for purists both in the disciplinary and epistemological sense. For example, I have found a great challenge in finding journals that suit the kind of work I am doing. I have had several editors tell me that, although the work was interesting, it focussed too much on disciplines other than their own. I have gone through countless editors trying to 'feel out' whether they would accept something that referenced education research in an article about policy-making and science communication. Similarly and perhaps unsurprisingly, I have also found qualitative research to be difficult to publish. The story completion paper has had an immense challenge in finding a home that is interested in social sciences methodology, data practices, and qualitative research. One memorable editor in fact told me there were 'no results presented' as they looked over the qualitative thematic analysis results. A not insignificant portion of my studies has been spent trying to grasp hold of where I stand in the academy, and more critically whether the work I am doing is of value to anyone but myself. When I faced writing up this thesis I found myself drawn towards justifying why this research is important as well as my personal positioning in addressing the research questions.

I also approached the PhD as a learning process, I did not think I needed to be an expert in every discipline of my interdisciplinary leanings. However, in publishing and discussing my thesis I find the academy does not always agree. Or rather that to get published I must hide the other disciplines I am using and write from the perspective of the journal at hand. One of the most practical results of this is the need to adapt to the use of terms like big data, data science, civic

datafication or data sharing dependent on the journal. This obscures the value of interdisciplinarity in helping create new models of thinking and find the similarities in disciplines that so often go unnoticed. Public engagement research, science communication, theories of governance, education research...they all talk about such similar things, yet few scholars have taken the time to look at how these disciplines fit together, nor who is repeating work that has been long established a few buildings over on campus. One must bluster through pretending to have discovered lightning when all they got was a static shock. Countering this is the value of interdisciplinarity.

This is not an indictment of the people who do occupy traditional academic disciplines, rather I am expressing that in writing up the work from a feminist perspective I may seem overly argumentative at times. I have felt the need to battle for my place as an emerging interdisciplinary scholar, as a mixed methods researcher, and as a woman in a male-dominated data world. I have needed to carve out a smaller and smaller space for myself so that I do not seem to be overstepping. Maybe this is a challenge that all researchers have faced at one time or another, but it nevertheless made completing this thesis challenging.

This thesis has been about pushing and pulling and discovering and discussing a broad range of topics. I am lucky enough to have supervisors who support this kind of winding study. As a closing thought I would like to say that due to the exploratory and winding nature of this thesis, this work is not perfect. I hope that it is challenged and debated and questioned. The main conclusion of this doctoral thesis is that I have learned, not that I have done perfect work, and the value of the PhD was in that learning.

11.2 Conclusion

This thesis aimed to explore how publics and policy-makers can engage with emerging data technologies in the context of local and national government in the UK. Utilising a mixed methods, primarily qualitative, approach this thesis found that publics are often limited in their inclusion both in the use of data technology in government and governance related to that use. Publics are

generally only thought of in relation to consent or in deference to the concept of 'public good' for policy decision-making. Despite the entrenched rhetoric around the necessary adoption of data technologies for government at every level, policy-makers have also not engaged strongly with data technologies. This is due both to resource limitations and to the narrative-making nature of government policy-making that does not always translate to the use of data technologies. Subsequently, this thesis proposes that publics should be included both in the practice and governance of data technology use through ongoing and long-term public engagement exercises. As government struggles with what ethical and good data practice consists of, there needs to be greater emphasis placed on public engagement models that work towards a collaborative imagination of the future of data technologies in the UK. Future research should focus on what this kind of public engagement looks like in practice, as well as how government can better regulate both their own use of data and the use of data in the private sector. Data technologies are imagined as both the saviour of complex social problems like poverty, as well as a dystopic premonition of government control. These imaginations evidence that data technologies require ongoing, critical, and impactful debate. In other words, data technologies need public engagement that values equally the technical and social expertise of different publics.

REFERENCES

Achen, C. H., & Bartels, L. M. (2016). Democracy for realists: Why elections do not produce responsive government. Princeton: Princeton University Press.

Adams, T. (2018). Facebook's week of shame: the Cambridge Analytica fallout. The Guardian. Retrieved from https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence

Andrews, E., Thornton, D., Owen, J., Bleasdale, A., Freeguard, G., & Stelk, I. (2016). Making a success of digital government. Retrieved from London, UK: https://www.instituteforgovernment.org.uk/sites/default/files/publications/IFGJ4942_Digital_Government_Report_10_16%20WEB%20%28a%29.pdf

Braun, K., & Konninger, S. (2017). From experiments to ecosystems? Reviewing public participation, scientific governance and the systemic turn. Public Understanding of Science. doi:10.1177/0963662517717375

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. doi:10.1191/1478088706qp063oa

Burgess, M. M. (2014). From 'trust us' to participatory governance: Deliberative publics and science policy. Public Understanding of Science, 23(1), 48-52. doi:10.1177/0963662512472160

Burri, R. V., & Bellucci, S. (2008). Public perception of nanotechnology. Journal of Nanoparticle Research, 10(3), 387-391. doi:10.1007/s11051-007-9286-7

Cabinet Office, & Government Digital Service. (2017). Government Transformation Strategy. London, UK.

Cairney, P., & Heikkila, T. (2014). A comparison of theories of the policy process. In P. Sabatier & C. Weible (Eds.), Theories of the Policy Process (3rd ed.). Chicago: Westview Press.

Cameron, D., Pope, S., Clemence, M., & Ipsos MORI Social Research Institute. (2014). Dialogue on Data - Exploring the public's views on using linked administrative data for research purposes. Retrieved from London, UK: https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care.data ran into trouble. Journal of Medical Ethics, 41(5), 404-409. doi:10.1136/medethics-2014-102374

Cotton, D., Winter, J., & Bailey, I. (2013). Researching the hidden curriculum: intentional and unintended messages. Journal of Geography in Higher Education, 37(2), 192-203. doi:10.1080/03098265.2012.733684

Dalrymple, T. (2001). The dystopian imagination. City Journal, Autumn 2001.

Data Justice Lab. (2018). Data Justice 2018 Programme. Paper presented at Data Justice 2018, Cardiff UK.

Data Power. (2017). Data Power 2017 Program: An international conference on (big) data & power. Paper presented at the Data Power 2017, Ottawa, ON.

Davidson, B. (2017). Storytelling and evidence-based policy: Lessons from the grey literature. Palgrave Communications, 3, 17093. doi:10.1057/palcomms.2017.93

Davidson, S., McLean, C., Treanor, S., Ipsos MORI, Aitken, M., Cunningham-Burley, S., . . . University of Edinburgh. (2013). Public acceptability of data sharing between the public, private and third sectors for research purposes. Retrieved from http://www.gov.scot/resource/0043/00435458.pdf

Davis, K. (2008). Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful. Feminist Theory, 9(1), 67-85.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., . . . Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment, 10(6), 291-297. doi:10.1890/110236

Farrell, D. M. (2014). 'Stripped Down' or Reconfigured Democracy. West European Politics, 37(2), 439-455. doi:10.1080/01402382.2014.887873

Felt, U., & Fochler, M. (2010). Machineries for Making Publics: Inscribing and De-scribing Publics in Public Engagement. Minerva, 48(3), 219-238. doi:10.1007/s11024-010-9155-x

Ferguson, K. E. (2017). Feminist theory today. Annual Review of Political Science, 20, 269-286. doi:10.1146/annurev-polisci-052715111648

Fiorino, D. J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. Science Technology Human Values, 15(2), 226-243.

Galston, W. A. (2010). Realism in political theory. European Journal of Political Theory, 9(4), 385-411. doi:10.1177/1474885110374001

Gill, S. (2017). A dialectic of utopia/dystopia in the public imagination of the 21st century. Global-E.

Ginnis, S., Evans, H., Boal, N., Davies, E., Palmqvist Aslaksen, A., Ipsos MORI Social Research Institute, . . . Sciencewise. (2016). Public dialogue on the ethics of data science in government. Retrieved from London, UK: https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf

Gov.UK. (2018). Services data. Retrieved from https://www.gov.uk/performance/services

Griggs, S., Norval, A. J., & Wagenaar, H. (2014). Introduction : Democracy, conflict and participation in decentred governance. In S. Griggs, A. J. Norval, & H. Wagenaar (Eds.), Practices of Freedom: Decentred Governance, Conflict and Democratic Participation. New York, NY: Cambridge University Press.

Groves, C. (2011). Public engagement and nanotechnology in the UK: Restoring trust or building robustness? Science and Public Policy, 38(10), 783-793. doi:10.3152/030234211x13070021633440

Hand, D. J. (2013). Data, not dogma: Big data, open data, and the opportunities ahead. Paper presented at Advances in Intelligent Data Analysis XII, London UK.

Hintz, A. (2018). Social trust, transformed into technological systems (such as blockchain) becomes antisocial trust. We're in a cris… https://t.co/1PvZ2BjbfA. Retrieved from https://twitter.com/arne_hz/status/998886998218870785

Internetlivestats.com. (2018). Twitter usage statistics. Retrieved from http://www.internetlivestats.com/twitter-statistics/

Ipsos MORI. (2006). The use of personal health information in medical research general public consultation: Final report. Retrieved from London, UK: https://mrc.ukri.org/documents/pdf/the-use-of-personal-health-information-in-medical-research-june-2007/

Ipsos MORI. (2016). The one-way mirror: Public attitudes to commercial access to health data. Retrieved from London, UK: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Janssen, M., & Helbig, N. (2016). Innovating and changing the policy-cycle: Policy-makers be prepared! Government Information Quarterly, Advance online publication. doi:10.1016/j.giq.2015.11.009

Jarrar, Y. (2017). What is the role of government in the digital age? World Economic Forum. Retrieved from https://www.weforum.org/agenda/2017/02/role-of-government-digital-age-data/

Johnson, J. A. (2014). From open data to information justice. Ethics and Information Technology, 16(4), 263-274. doi:10.1007/s10676-014-9351-8

Keller, S. A., Koonin, S. E., & Shipp, S. (2012). Big data and city living – what can it do for us? Significance, 9(4), 4-7.

Kettl, D. F. (2016). Making data speak: Lessons for using numbers for solving public policy puzzles. Governance, 29(4), 573-579. doi:10.1111/gove.12211

Kitchin, R. (2013). The real-time city? Big data and smart urbanism. GeoJournal, 79(1), 1-14. doi:10.1007/s10708-013-9516-8

Lauriault, T. P., & Mooney, P. (2014). Crowdsourcing: A geographic approach to public engagement. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2518233

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public engagement. Government Information Quarterly, 29(4), 492-503. doi:10.1016/j.giq.2012.06.001

Lepore, J. (2017). A golden age for dystopian fiction. The New Yorker.

Longo, J., Kuras, E., Smith, H., Hondula, D. M., & Johnston, E. (2017). Technology Use, Exposure to Natural Hazards, and Being Digitally Invisible: Implications for Policy Analytics. Policy and Internet, 9(1), 76-108. doi:10.1002/poi3.144

Malomo, F., & Sena, V. (2017). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. Policy & Internet, 9(1), 7-27.

Manzoni, J. (2017). Speech: Big data in government: the challenges and opportunities. Retrieved from:

https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities

Markus, M. L., & Topi, H. (2015). Big data, big decisions for science, society, and business: Report on a research agenda setting workshop. Washington, D.C. Retrieved from https://dl.acm.org/citation.cfm?id=2849516

McNeely, C. L., & Hamh, J. (2014). The big (data) bang: policy, prospects, and challenges. Review of Policy Research, 31(4), 304-310.

Michael, M., & Lupton, D. (2016). Toward a manifesto for the 'public understanding of big data'. Public Understanding of Science, 25(1), 104-116. doi:10.1177/0963662515609005

Milakovich, M. E. (2013). Anticipatory government: Integrating big data for smaller government. Paper presented at the Internet, Politics, Policy 2012, Oxford.

Mjøset, L. (2009). The contextualist approach to social science methodology. In D. Byrne & C. C. Ragin (Eds.), The SAGE Handbook of Case-Based Methods (pp. 39-68). London, UK: SAGE Publications.

Newman, J. (2011). Public leadership as public-making. Public Money & Management, 31(5), 315-322. doi:10.1080/09540962.2011.598336

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. International Journal of Social Research Methodology, 8(5), 375-387. doi:10.1080/13645570500402447

Orr, K., & Vince, R. (2009). Traditions of local government. Public Administration, 87(3), 655-677. doi:10.1111/j.1467-9299.2009.01770.x

Parliament.UK. (2018). Parliament and the government. Retrieved from https://www.parliament.uk/about/how/role/parliament-government/

Piscopo, A., Siebes, R., & Hardman, L. (2017). Predicting sense of community and participation by applying machine learning to open government data. Policy and Internet, 9(1), 55-75. doi:10.1002/poi3.145

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. Health and Technology. doi:10.1007/s12553-017-0179-1

Rempel, E. S. (2015). Intersecting Risk Factors: Diagnosed alcohol dependence and criminal sentencing in British Columbia's Aboriginal Populations. (Masters), Simon Fraser University, Burnaby, Canada.

Rempel, E. S., Somers, J. M., Calvert, J. R., & McCandless, L. C. (2015). Diagnosed alcohol dependence and criminal sentencing among British Columbian Aboriginal offenders. Drug and Alcohol Dependence, 154, 192-198.

Rempel, E. S., Wilson, E. N., Durrant, H., & Barnett, J. (2017). Preparing the prescription: a review of the aim and measurement of social referral programmes. BMJ Open, 7. doi:10.1136/bmjopen-2017-017734

Renn, O. (2008). Risk Participation. In O. Renn (Ed.), Risk Governance: Coping with Uncertainty in a Complex World. London: Earthscan.

Riesch, H., Potter, C., & Davies, L. (2013). Combining citizen science and public engagement: The Open AirLaboratories Programme. Journal of Science Communication, 12(3), 1-19.

Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. Science, Technology & Human Values, 30(2), 251-290. doi:10.1177/0162243904271724

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. New Media & Society, 18(4), 581-599.

Selin, C., Rawlings, K. C., de Ridder-Vignone, K., Sadowski, J., Altamirano Allende, C., Gano, G., . . . Guston, D. H. (2017). Experiments in engagement: Designing public engagement with science and technology for capacity building. Public Understanding of Science, 26(6), 634-649. doi:10.1177/0963662515620970

Setälä, M. (2017). Connecting deliberative mini-publics to representative decision making. European Journal of Political Research, 56(4), 846-863. doi:10.1111/1475-6765.12207

Spielberg, S. (Writer). (2002). Minority Report. In B. Curtis & J. de Bont (Producer). USA: Twentieth Century Fox.

Stilgoe, J., Irwin, A., & Jones, K. (2006). The received wisdom: Opening up expert advice. Retrieved from London, UK: https://www.demos.co.uk/files/receivedwisdom.pdf

Stilgoe, J., Lock, S. J., & Wilsdon, J. (2014). Why should we promote public engagement with science? Public Understanding of Science, 23(1), 4-15.

Stone, D. A. (1989). Causal stories and the formation of policy agendas. Political Science Quarterly, 104(2), 281-300.

Teare, H. J., Morrison, M., Whitley, E. A., & Kaye, J. (2015). Towards 'Engagement 2.0': Insights from a study of dynamic consent with biobank participants. Digit Health, 1, 2055207615605644. doi:10.1177/2055207615605644

Techopedia Inc. (2018). Datafication. Retrieved from https://www.techopedia.com/definition/30203/datafication

The Royal Society, British Academy, Fourniol, F., & McLaughlin, F. (2017). Data governance: Public engagement review. Retrieved from London, UK: https://royalsociety.org/~/media/policy/projects/data-governance/data-governance-public-engagement-review.pdf

UK House of Lords. (2000). Select committee on science and technology: Science and society. London, UK.

Vasileiou, K. (2015). Symbolic and practical facets in the use and production of home medical technology: The example of blood pressure monitoring (Doctoral dissertation). Retrieved from: https://bura.brunel.ac.uk/bitstream/2438/13821/1/FulltextThesis.pdf

Walls, J., O'Riordan, T., Horlick-Jones, T., & Niewohner, J. (2005). The meta-governance of risk and new technologies: GM crops and mobile telephones. Journal of Risk Research, 8(7-8), 635-661. doi:10.1080/13669870500101200

Ward, J. S., & Barker, A. (2013). Undefined by data: A survey of big data definition. University of St Andrews. St Andrews, UK.

Appendix A: Research Tools

The research tools include the story completion exercise designed for Chapter 5, the interview schedule used in Chapter 6, the workshop schedule used in Chapter 7 and lastly the survey instruments used in Chapter 8. They are listed below in that order.

CHAPTER 5 STORY COMPLETION EXERCISE

Emily S. Rempel
Department of Psychology
Institute for Policy Research
University of Bath

Story Completion Exercise
Project Title: "New frontiers of data in local UK government."
Please read the following prompt carefully and complete the story. It can be as long or short as you like, feel free to take extra space on the back of this form if you need it. There are no wrong answers. Thank you!
Sam is a local community organisation manager. She/He wants to understand how to get access to local council data about the people who use her/his organisation's services. She/He heard that there were local people in the council who she/he could contact. What does Sam do next?

CHAPTER 6 INTERVIEW SCHEDULE

Emily S. Rempel
Department of Psychology
Institute for Policy Research
University of Bath

Semi-structured Interview Schedule
Project Title: "New frontiers of data in local UK government."

I. Opening
   a. Introduction
      i. Introduce yourself, your role as a PhD student at the University of Bath and your overall research project.
      ii. "Today, I will be asking you some questions about your recent involvement with _____ (name of data project). The information you provide today will feed in to our understanding of the potential for data in the policy process."
      iii. "The interview will last at most 60-90 minutes and thank you in advance for agreeing to take part."
      iv. Introduce consent forms and allow time to read: "As part of this project, I will be recording this interview and taking notes. All of your answers will be kept confidential among the project staff, however I may use direct quotes in published materials. Any reference to what you say in this interview, will be anonymous. Taking part is voluntary and you may skip any questions. If you decide not to take part this will not impact your relationship with the University of Bath and you may withdraw for any reason, at any time. Prior to publication, the B&NES research manager, Jon Poole, will review the final report which will include anonymised quotes."
      v. Ask for questions then answer and if no questions, proceed to signing the consent form.

II. Transition (Now we will move on to the main questions for today.)

III. Topics
   a. Take me through your role in this project.
      i. What was the purpose of the project?
      ii. Did you use 'data'? If so, what kind of data did you use?
         1. What do you consider do be data?
      iii.
      iv. What kind of resources did you use in your work?
         1. Documents?
         2. Materials?
         3. Computing resources?
         4. Can you show me?
      v. Did you feel the project met its aim?

b. What kind of other projects have you heard of using 'data'?
    i. What do you consider to be 'new forms of data'?
    ii. What is the role of 'data in policy'?
    iii. How do you imagine the future of data use at the council?
c. What kind of people did you interact with?
    i. Do you feel comfortable providing me with their names?
    ii. Is it alright if I contact them, referencing you?

IV. Transition (Thank you for your answers, I am going to now summarize what we discussed to ensure that I have accurately interpreted our session.)

V. Closing
a. Summarize the main points from the previous questions.
b. Thank the interviewee for their time, remind them of the purpose of the interview and that they are free to email any questions they have in the future.
c. Close by reminding the interviewee that the results will be disseminated once the report is completed and they will be kept up to date on the project going forward.

CHAPTER 7 WORKSHOP SCHEDULE

Co-designing a Data Science Engagement Tool
Workshop Schedule

Introductions (5 MINUTES)

Part 1: *Ethics and Data Protection* (20 MINUTES) – TABLE MODERATORS

First: Pass out consent forms and ask participants to review and fill in.

1) Prompt 1: What do you think of filling in these kinds of forms?
2) Prompt 2: On the consent form there was a bit that said that your data will be kept 'confidential' and 'anonymised', what does that mean to you?
3) Prompt 3: Are you aware of any times that local government has asked for your consent?
    a. What about your GP?
    b. What about any websites like Facebook or Google?

Part 2: *Learning about data, data science, government, and engagement* (30 MINUTES) – TABLE MODERATORS

1) Data Activity/Presentation - EMILY REMPEL
2) Discussion
    a. Prompt 1: What are some examples of how you produce data day-to-day?
    b. Prompt 2: How do you think local government uses data for?
    c. Prompt 3: What sorts of things worry you about the sorts of things that local government use data for?
    d. Prompt 4: What sorts of things could/should data do to make our communities better?

Part 3: Potential for *Filling out the efficacy and engagement survey* (8-10 MINUTES)

----------BREAK FOR 20 MINUTES----------

Part 4: *Co-redesign of the online engagement platform* (45 MINUTES) – TABLE MODERATORS

1) Introduce prototype to participants.
2) Ask them to look it over for 10 minutes.
3) Ask them to read over and annotate while discussing.
    a. Prompt 1: What strikes you about what would and wouldn't work on a platform like this?
        i. Reminder: Platform aim is to develop collaborative data projects that citizens have a voice in.
    b. Prompt 2: How likely do you think you would be to use something like this?
        i. Why/Why not?
    c. Prompt 3: What sorts of things would make the platform work better for you?

Part 5: Wrap-up and workshop feedback (5 MINUTES)

----------END----------

CHAPTER 7 WORKSHOP FEEDBACK FORM

Workshop Feedback

Considering the workshop you just participated in please rank how strongly you agree with the following statement.

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
I enjoyed the event	○	○	○	○	○	○	○

Which part did you enjoy the most?

_____

_____

_____

_____

_____

Which part did you enjoy the least?

_____

_____

_____

_____

_____

What do you think could be changed about the workshop?

_____

_____

_____

_____

_____

_____

Were you able to express your views freely and openly?

- ○ Yes, completely
- ○ Yes, but sometimes I felt nervous
- ○ Not as much as I would have liked
- ○ Not at all
- ○ Don't know

Please use this space to add any additional comments on the workshop:

_____

_____

_____

_____

_____

_____

Q13 Please enter your year of birth:

_____

Q14 Please tell us your current employment status:

- ○ Full time education (more than 16 hours per week) (1)
- ○ Part time education (less than 16 hours per week) (2)
- ○ Apprenticeship/Internship (3)
- ○ Working full time (more than 16 hours per week) (4)
- ○ Working part time (less than 16 hours per week) (5)
- ○ Stay at home parent (7)
- ○ Volunteering (8)
- ○ Unemployed (9)
- ○ Retired (10)
- ○ Prefer not to say (11)

Q15 Please tell us your gender:

- ○ Male (1)
- ○ Female (2)
- ○ Not listed, please write in: (3) _____
- ○ Prefer not to say (4)

Q16 Please select your ethnicity category:

- ○ White (1)
- ○ Mixed Heritage - White and Black Caribbean (6)
- ○ Mixed Heritage - White and Black African (2)
- ○ Mixed Heritage - White and Asian (7)
- ○ Mixed Heritage - Other (3)
- ○ Asian - Chinese (8)
- ○ Asian - Indian (4)
- ○ Asian - Pakistani (9)
- ○ Asian - Bangladeshi (5)
- ○ Asian - Other (10)
- ○ Black - Caribbean (11)
- ○ Black - African (12)
- ○ Black - Other (13)
- ○ Any other ethnic background, please specify (14)
  _____
- ○ Prefer not to say (15)

Q17 Please select your highest level of education:

- GCSE/O-Level/CSE/Scottish Standards or Equivalent (1)
- Vocational qualifications (2)
- A-level/Scottish Highers or Equivalent (3)
- Bachelor Degree or Equivalent (4)
- Masters/PhD or Equivalent (5)
- Other (6)
- No formal qualifications (7)
- Don't know (8)
- Prefer not to say (9)

Q18 How much, if anything, do you feel you know about how local government works?

- A great deal (1)
- A fair amount (2)
- Not very much (3)
- Nothing at all (4)

Q19 How much influence, if any, do you feel you have over decision making in your local area?

- A great deal (1)
- A fair amount (2)
- Not very much (3)
- Nothing at all (4)

Q20 How much do you believe that your local officials care about what you think is important for your local area?

- Very much (1)
- Somewhat (2)
- Only a little (3)
- Not at all (4)

Q21 How confident are you in your own ability to participate in local politics?

- Not at all confident (1)
- A little confident (2)
- Quite confident (3)
- Very confident (4)
- Completely confident (5)
- Don't know (6)

Q22 To what extent do you agree or disagree with the following statement:
When people like me get involved in politics, they really can change the way their local area is run?

- Strongly agree (1)
- Tend to agree (2)
- Neither agree/nor disagree (3)
- Tend to disagree (4)
- Strongly disagree (5)

Q23 How satisfied are you with the way local government in your area is doing its job?

- Very satisfied (1)
- Fairly satisfied (2)
- Fairly dissatisfied (3)
- Very dissatisfied (4)

Q24 Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?

Please give a score between 0 to 10, where 0 means you strongly agree that you can't be too careful and 10 means that you strongly agree that most people can be trusted.

	Can't be too careful									Most people can be trusted
	0	1	2	3	4	5	6	7	8	9	10
Please drag the bar: (1)											

Q47
To what extent do you agree or disagree with the following statement:
Modern data science is so complex that public involvement is not realistic.

- Strongly agree (1)
- Somewhat agree (2)
- Neither agree nor disagree (3)
- Somewhat disagree (4)
- Strongly disagree (5)

Q25 We are now going to ask you some questions about your interest in politics:

	Not at all interested (1)	Not very interested (2)	Fairly Interested (3)	Very interested (4)
How interested would you say you are in national politics? (1)	○	○	○	○
How interested would you say you are in local politics? (2)	○	○	○	○

Q26 Did you vote in the:

	Yes (1)	No (2)	Don't Know (3)	Wasn't Eligible (4)
2017 General Election (1)	○	○	○	○
2016 EU Referendum (2)	○	○	○	○

Q27 In the last 12 months have you done any of the following to influence decisions, laws or policies?

	No (1)	Maybe (2)	Yes (3)
Donated money or pay a membership fee to a charity or campaigning organisation (1)	○	○	○
Created or signed an e-petition (2)	○	○	○
Created or signed a paper petition (3)	○	○	○
Contacted a local councillor or MP (4)	○	○	○
Boycotted certain products for political, ethical or environmental reasons (5)	○	○	○
Contributed to a discussion or campaign online or on social media (6)	○	○	○
Taken part in a public consultation (7)	○	○	○
Taken an active part in a campaign (8)	○	○	○
Attended political meetings (9)	○	○	○
Donated money or paid a membership fee to a political party (10)	○	○	○
Taken part in a demonstration, picket or march (11)	○	○	○
Contacted the media (12)	○	○	○
Submitted an Freedom of Information request (13)	○	○	○
Voted in an election (14)	○	○	○

Q28 How important is each of the following in being a good citizen?

	Very important (1)	Fairly important (2)	Neither important nor unimportant (3)	Fairly unimportant (4)	Very unimportant (5)
Obeying the law (1)	○	○	○	○	○
Participating in activities to benefit people in the community (2)	○	○	○	○	○
Taking part in activities to protect the environment (3)	○	○	○	○	○
Voting in elections (4)	○	○	○	○	○
Taking part in activities promoting human rights (5)	○	○	○	○	○
Talking with your family and friends about political issues (6)	○	○	○	○	○
Following political issues in the news (7)	○	○	○	○	○
Participating in a peaceful protest against a law you believe unjust (8)	○	○	○	○	○
Knowing about the country's history (9)	○	○	○	○	○

Joining a political party (10)	○	○	○	○	○	

Q29 Now we are going to ask you some questions about your digital skills:

	Not at all true of me (1)	Not very true of me (2)	Neither true nor untrue of me (3)	Mostly true of me (4)	Very true of me (5)	I do not understand what you mean by that (6)
I know which information I should and shouldn't share online (1)	○	○	○	○	○	○
I know when I should and shouldn't share information online (2)	○	○	○	○	○	○
I am careful to make my comments and behaviours appropriate to the situation I find myself in online (3)	○	○	○	○	○	○
I know how to change who I share content with (e.g. friends, friends of friends or public) (4)	○	○	○	○	○	○
I know how to remove friends from my contact lists (5)	○	○	○	○	○	○

Q30 Please select the relevant statement which you feel applies to your competence in computer programming:

- I don't know what programming is. (1)
- I am aware of what programming is but cannot use a programming language. (2)
- I know the basics of one programming language. (3)
- I can use one or more programming languages to design, create and modify databases with a computer tool. (4)

Q48 On the next page you will be shown a type of data or information, for example your health records or your grocery purchases. Please take a note of what type of data you are presented. The final two pages will ask you to answer questions related to that type of data.

If you feel you do not create this kind of data, for example you don't have a mobile phone, please answer as if you did. You can press the back button at any time if you forget your type of data.

Q37 Your posts on public social media sites like Twitter, Reddit, or TripAdvisor.

Q38 Your location (GPS) information from your mobile phone.

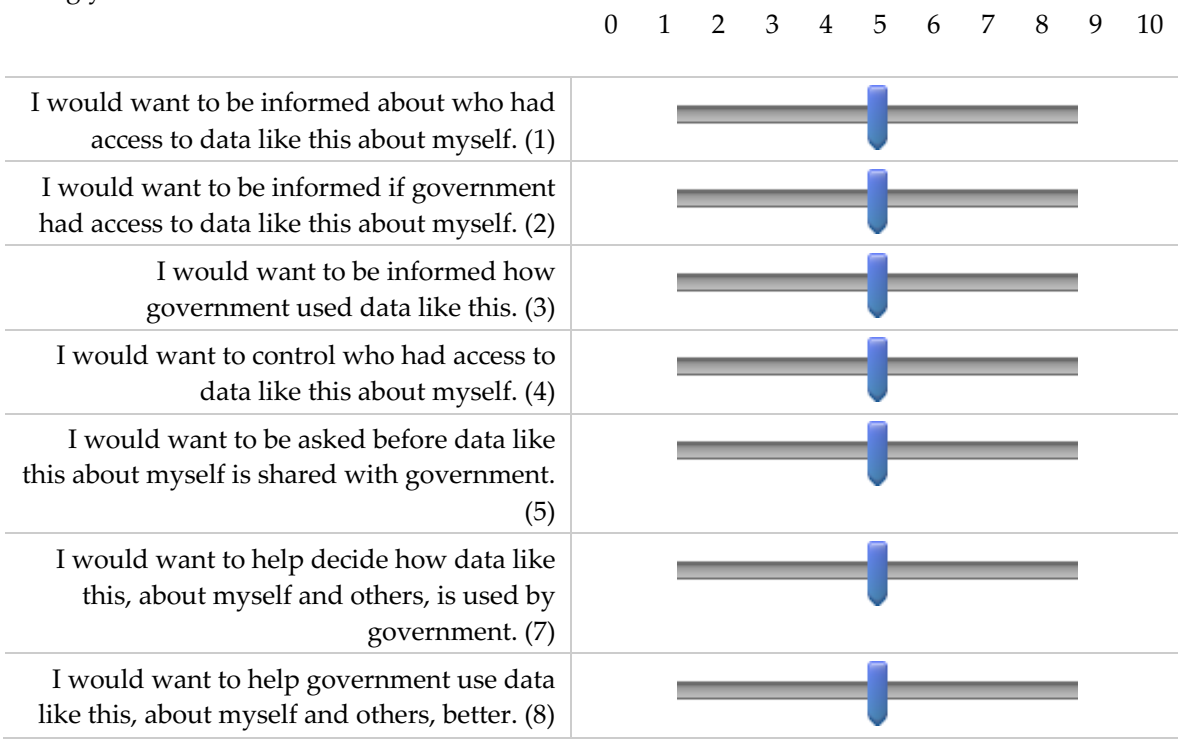
Q39 Your health records from your GP.

Q40 Your purchase history from a store card or online store.

Q41 Your employment history.

Q42 When considering this kind of information/data that is created by you and other people you know, please rank below how strongly you feel about each of the following statements from 0 to 10 where 0 is not at all strongly, and 10 is very strongly.

Q43 When considering this kind of information/data, please rank below how strongly you feel about each of the following statements from 0 to 10 where 0 is not at all strongly, and 10 is very strongly.



Q44

Q45 Thank you for taking part in the survey! Here are our contact details if you'd like to get in touch. If you would like to receive a summary of the results of the survey please feel free to email us and we will send you the results once they are available.

Emily Rempel
Department of Psychology, University of Bath
Bath UK BA2 7AY
Email: E.S.Rempel@bath.ac.uk Phone: 01225 384573

Appendix B: Ethics Approval

Included in Appendix B are the final research ethics approval emails provided by the University of Bath's Department of Psychology Research Ethics Board. As well, the blank consent forms are included for each chapter. All projects were approved and their individual codes and confirmation are listed below. Chapter 3 does not have a research ethics statement as it was a critical literature review and did not involve human participants or the collection of personal information.

From: psychology-ethics
Sent: 21 June 2016 13:38
To: Emily Rempel <E.S.Rempel@bath.ac.uk>
Cc: Hannah Durrant <H.Durrant@bath.ac.uk>; Julie Barnett <J.C.Barnett@bath.ac.uk>
Subject: RE: Ethics 16-136

Dear Emily Rempel

Reference Number 16-136: New frontiers of data in local UK government

Thank you for satisfactorily attending to those amendments. I can now confirm that you have full ethical approval for your study.

Best wishes with your research,

Dr Michael J Proulx

Chair, Psychology Research Ethics Committee

Memorandum of Understanding

Project Title: "Public engagement and policy-making in the era of big data."

Between:

Emily Rempel

PhD Student

Department of Psychology

University of Bath

And

[REDACTED]

Background

As part of a PhD dissertation, this study will examine how big data and data science are currently conceptualised in the policy-making process at the National UK level. As well as how public engagement is operationalised in data-driven policy making. Emily Rempel, as a researcher at the University of Bath, will collect ethnographic data as a participant-observer during the workshop planning and analysis of the UK Government's Public Dialogue on Data Science Ethics. This will include note taking at meetings and during the workshop activities. No personal or sensitive organizational data will be collected from any individual during the research process. Further interviews with key stakeholders will occur and consent will be sought on a case-by-case basis.

Purpose

To provide an agreement for ethnographic work to take place in relation to the Public Dialogue on Data Science Ethics workshop and meetings, as part of the Government Digital Service.

Reporting

Data from the ethnographic work will be collated and analysed for the purpose of academic publication. No direct quotes from the ethnography will be used and all individuals will be anonymous. Prior to submission for publication, the work will be submitted to Cat Drew for review. Any information deemed inappropriate or sensitive will be redacted.

Duration

Data collection will occur from January 2016 to June 2016.

Name _____

Position _____

Organization _____

Date _____

Sign _____

Name Emily Rempel

Position PhD Researcher

Organization University of Bath

Date _____

Sign _____

CHAPTER 5 AND 6 RESEARCH ETHICS APPROVAL, INFORMATION SHEETS, AND CONSENT FORMS

From: psychology-ethics <psychology-ethics@bath.ac.uk>
Date: March 11, 2016 at 1:35:45 PM GMT
To: Emily Rempel <E.S.Rempel@bath.ac.uk>
Subject: RE: Ethics 16-044

Dear Emily Rempel

Reference Number 16-044: Public engagement and policy-making in the era of big data

Thank you for satisfactorily attending to those amendments. I can now confirm that you have full ethical approval for your study.

Best wishes with your research,

Dr Michael J Proulx

Chair, Psychology Research Ethics Committee

Consent to Participate in a Personal Interview Information Sheet

Project Title: "New frontiers of data in local UK government."

Quick Summary:

As part of a PhD dissertation within the Department of Psychology, this study will examine how new forms of data are used in policy-making in a local UK government setting.

What is this study about?

New forms of data like 'big data' and 'open data' are increasingly common buzz words for the future of policy development. However, there is limited evidence on how new kinds of data are used in policy. It is essential to have a baseline understanding of how government and policymakers interact with data, to realise this potential of a more data-driven policy future. This project aims to identify current and theoretical uses for new forms of data in local policy development. New forms of data include things like social media data, 'big' company data and smart meter readings. We will do this using semi-structured interviews with individuals who have taken part in or commissioned the use of data projects within Bath and North East Somerset Council.

What we will ask you to do:

This interview will be one-on-one and will last between 60-90 minutes. You will be asked several questions about your experiences with data projects. As well, you will be asked to complete a hypothetical story about the use of data in local UK government. You will also be asked some personal information including your age and gender. Your interview will be audio recorded.

Risks and benefits:

Your personal data and answers will be kept entirely confidential. Your interview data will be anonymised and no identifiable information will be linked to you. However, the information you provide including direct quotes will be used in published material by the researchers. In addition, the Research & Intelligence Manager at Bath and North East Somerset Council will be provided a copy of all reports, which will include anonymised quotes, prior to publication. Taking part is voluntary and you may skip any questions. If you decide to not take part, this will not affect your relationship with the University of Bath. You may withdraw, without reason, at any time.

Sharing results:

The final results of these interviews will be analysed and submitted in dissertation format as well as in academic publications. Prior to publication, the results will be disseminated to all research participants.

If you have any questions, please do not hesitate to ask them now. Alternatively, further questions may be addressed to:

Emily Rempel (Department of Psychology, University of Bath, Bath UK BA27AY)

Email: e.s.rempel@bath.ac.uk Phone: [redacted]

Informed Consent of Interview Participants

Project Title: "New frontiers of data in local UK government."

- ❑ I confirm that I have read and understood the information about the study provided on the information sheet and have had time to ask questions. I confirm that these questions have been answered to my satisfaction.
- ❑ I consent voluntarily to participate in the study as outlined above and understand that I can withdraw at any time.
- ❑ I consent to participate in a structured interview, which will be audio recorded.
- ❑ I consent to write and submit a 'story completion exercise' for this study.
- ❑ I consent to the use of anonymised quotes from the interview and story exercise in publication.

Participant Name _____ **Date**_____

Signature_____

Interviewer Name _____ **Date**_____

Signature_____

Contact Information:

Emily Rempel

Department of Psychology, University of Bath

Bath UK BA27AY

e.s.rempel@bath.ac.uk

[redacted]

CHAPTER 7 RESEARCH ETHICS APPROVAL AND CONSENT FORM

From: psychology-ethics
Sent: Wednesday, December 13, 2017 2:47 PM
To: Emily Rempel
Cc: Julie Barnett
Subject: Ethics 17-315 Approved

Dear Emily,

Ethics code 17-315 Collaborating to Co-design a Data Science Engagement Tool

I am happy to confirm that you have received full ethical approval from the University of Bath Department of Psychology Ethics Committee for your application. In light of the fact that this project has no serious ethical concerns, this approval has been granted via Chair's Action. Please use the code 17-315 as proof of ethical approval on internal documents and any subsequent emails.

Best of luck with your research,

Dr. Nathalia Gjersoe

Chair, Psychology Ethics Committee

CONSENT FORM
Co-designing a Data Science Engagement Tool

Please answer the following questions to the best of your knowledge

	YES	NO
DO YOU CONFIRM THAT YOU:		
• Are over 18 years of age.		
HAVE YOU:		
• Been given information explaining about the study?		
• Had an opportunity to ask questions and discuss this study?		
• Received satisfactory answers to all questions you asked?		
• Received enough information about the study for you to make a decision about your participation?		
DO YOU UNDERSTAND:		
that you are free to withdraw your SURVEY data, but NOT AUDIO OR WORKSHOP DATA, prior to anonymization:		
• At any time?		
• Without having to give a reason for withdrawing?		

I hereby fully and freely consent to my participation in this study

I understand the nature and purpose of the procedures involved in this study. These have been communicated to me on the information sheet accompanying this form.

I understand and acknowledge that the investigation is designed to promote scientific knowledge and that the University of Bath will use the data I provide for no purpose other than research.

I understand that the data I provide will be kept **confidential**, and that on completion of the study my data will be **anonymised** by removing all links between my name or other identifying information and my study data. This will be done by March 1, 2018, and before any presentation or publication of my data.

I understand that after the study the data will be made "open data". I understand that this means the anonymised data, along with the results of the workshop, will be publicly available and may be used for purposes not related to this study, and it will not be possible to identify me from these data.

Participant's signature: _____ Date:

_____

Name in BLOCK Letters: _____

Final consent

Having participated in this study

I agree to the University of Bath keeping and processing the data I have provided during the course of this study in accordance with the information I received at the outset.

Participant's signature: _____ Date:

_____

Name in BLOCK Letters: _____

If you have any concerns related to your participation in this study please direct them to the Department of Psychology Research Ethics Committee, via Nathalia Gjersoe Research Ethics Officer (Tel: 01225 38 3251 email: N.Gjersoe@bath.ac.uk).

From: psychology-ethics
Sent: Monday, March 5, 2018 1:02 PM
To: Emily Rempel
Cc: Julie Barnett
Subject: Ethics 18-037 Approved

Dear Emily,

Ethics code 18-037 Views on local government, data use, and public influence

I am happy to confirm that you have received full ethical approval from the University of Bath Department of Psychology Ethics Committee for your application. In light of the fact that this project has no serious ethical implications, this approval has been granted via Chair's Action. Please use the code18-037 as proof of ethical approval on internal documentation.

Please be aware that it is your responsibility to let us know as soon as possible if any issues or complaints of an ethical nature arise over the course of your research.

Best of luck with your research,

Dr. Nathalia Gjersoe

Chair, Psychology Ethics Committee

Views on local government and data use

Q2 *Survey Information Project Title: "Co-designing a data science engagement tool."*

Quick Summary: As part of a PhD thesis within the Department of Psychology, we want to look at what the public thinks about government data use.

What is this study about? Data science, which is about combining and using data in different ways, is an increasingly popular technology. Although you may feel like you don't know much about it, you interact with data and data science every day. For example, when you see an ad on a website for something you searched Google for yesterday. Or even when you get personalised coupons to save money on products you buy most often. Data and information, often survey data or local information about your area, can also be used by local government to help communities. But individuals don't often have influence over how their data is used or what it is used for. We want to understand how individuals and publics want to be involved, or engaged, in the use of their data.

What we will ask you to do: We want you to help us by answering a few questions about your views on government, your opinion on different kinds of data (e.g. Facebook information versus employment records), and how you would want to be involved in the use of that kind of data. To help us understand how these views are different with different groups of people we will also ask you a bit about your age and background – we don't need to know anything like your name or address though.

Risks and benefits: Your personal data and answers will be kept entirely confidential. Your survey data will be anonymised and no identifiable information will be linked to you. We will publish the final anonymised data online as 'open data'. We are doing this because we feel an important part of scientific research is transparency and openness. Therefore other researchers may use this data in the future. Taking part is voluntary and you may stop the survey at any time you wish.

Sharing results:
The final results of this survey will be analysed and presented in a PhD thesis as well as in academic publications

Q1

Q3 Please enter your Prolific ID:

_____

Q4 If you have any questions before taking part in the survey please contact us at:
Emily Rempel Department of Psychology, University of Bath Bath UK BA27AY
e.s.rempel@bath.ac.uk 01225.384573

Q5

Q6 Informed Consent to Use Personal Information
Please return your submission to Prolific if you do not wish to click 'yes' to the following consent statements. If you click no, your survey session will end. Thank you!

Q7 I confirm that I have read and understood the information about the study provided on the previous page, and if needed I have contacted the research team to ask any questions I may have. I confirm that these questions, if asked, have been answered acceptably.

- ○ Yes (1)
- ○ No (2)

Q8
I understand that after the study the data will be made “open”. I understand that this means the anonymised data, will be publicly available and may be used for purposes not related to this study, and it will not be possible to identify me from these data.

- ○ Yes (1)
- ○ No (2)

Q9 I consent to the use of my information in write-ups and summaries of this project.

- ○ Yes (1)
- ○ No (2)

Q10 I confirm that I am over the age of 18.

- ○ Yes (1)
- ○ No (2)

Q11 I understand that I may withdraw my consent at any time, without reason, but that once my response is anonymised I cannot withdraw my data.

- ○ Yes (1)
- ○ No (2)

Q12 If you'd like, please print a copy of this consent form for your records! I have read and understand the above consent form, by clicking the submit button to enter my information, I indicate my willingness to voluntarily provide my consent as outlined above.

Appendix C: Additional Published Works

Appendix C presents two additional published works related to the aim of the thesis, however as these works were not completed under the thesis funding but as part of additional research undertaken by the author they are not presented as empirical chapters. They are listed here, along with authorship declarations, and discussed in Chapter 9: Additional Research.

ARTICLE: PREPARING THE PRESCRIPTION: A REVIEW OF THE AIM AND MEASUREMENT OF SOCIAL REFERRAL PROGRAMMES.

This declaration concerns the article entitled:									
Preparing the prescription: a review of the aim and measurement of social referral programmes.									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**		**Accepted**		**Published**	✓
Publication details (reference)	Rempel E.S., Wilson E.N., Durrant H., and Barnett J. (2017). Preparing the prescription: A review of the aim and measurement of social referral programmes. *BMJ Open, 7(e018834),* doi: 10.1136/bmjopen-2017-017734								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate considerably contributed to the formulation of ideas and methodology design, work, and presentation of data. Formulation of ideas: ESR was responsible for 50% of the formulation of ideas while HD was responsible for 40% and JB 10%. Design of methodology: ESR was responsible for 70% of the methodology literature review strategy with HD responsible for the other 30%. Experimental work: ESR conducted the literature review (60%) with EN as the second coder (30%) and with review from JB (10%). Presentation of data in journal format: ESR wrote up the results in journal format (60%) with JB also writing and editing (40%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							**Date**	01/06/2018	

Open Access Research

BMJ Open

Preparing the prescription: a review of the aim and measurement of social referral programmes

Emily S Rempel,[1] Emma N Wilson,[1] Hannah Durrant,[2] Julie Barnett[1,2]

To cite: Rempel ES, Wilson EN, Durrant H, *et al*. Preparing the prescription: a review of the aim and measurement of social referral programmes. *BMJ Open* 2017;**7**:e017734. doi:10.1136/bmjopen-2017-017734

► Prepublication history and additional material for this paper are available online. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2017-017734).

Received 15 May 2017
Revised 15 July 2017
Accepted 3 August 2017

[1]Department of Psychology, University of Bath, Bath, UK
[2]Institute for Policy Research, University of Bath, Bath, UK

Correspondence to
Emily S Rempel;
e.s.rempel@bath.ac.uk

ABSTRACT

Objective Our aim is to review, and qualitatively evaluate, the aims and measures of social referral programmes. Our first objective is to identify the aims of social referral initiatives. Our second objective is to identify the measures used to evaluate whether the aims of social referral were met.

Design Literature review.

Background Social referral programmes, also called social prescribing and emergency case referral, link primary and secondary healthcare with community services, often under the guise of decreasing health system costs.

Method Following the PRISMA guidelines, we undertook a literature review to address that aim. We searched in five academic online databases and in one online non-academic search engine, including both academic and grey literature, for articles referring to 'social prescribing' or 'community referral'.

Results We identified 41 relevant articles and reports. After extracting the aims, measures and type of study, we found that most social referral programmes aimed to address a wide variety of system and individual health problems. This included cost savings, resource reallocation and improved mental, physical and social well-being. Across the 41 studies and reports, there were 154 different kinds of measures or methods of evaluation identified. Of these, the most commonly used individual measure was the Warwick-Edinburgh Mental Well-being Scale, used in nine studies and reports.

Conclusions These inconsistencies in aims and measures used pose serious problems when social prescribing and other referral programmes are often advertised as a solution to health services-budgeting constraints, as well as a range of chronic mental and physical health conditions. We recommend researchers and local community organisers alike to critically evaluate for whom, where and why their social referral programmes 'work'.

Strengths and limitations of this study

- A strength of this study was the inclusion of both grey and academic literature to ensure a broad representation of social referral programmes.
- A strength of this study is in the review of aims and measures of social referral programmes, rather than outcomes.
- A limitation of this study was that there is no guarantee of an entirely comprehensive inclusion of all relevant articles; for example, we only accessed articles and reports available online or through the British Library.
- A limitation of this study was the use of the search term 'social prescribing' as this is a generalised UK region-specific term; however, this is the term used colloquially to describe social referral programmes.

INTRODUCTION

"The tonic effect of fun and play has long been recognized as an antidote to the stresses, worries, labors, and responsibilities of our workaday life…we must diagnose and prepare the prescription."[1] In 1958, Walt Disney wrote this commentary on film and American life for the 75th anniversary of the *Journal of the American Medical Association*. Although few would argue that Disney was a great early adopter of the social determinants of health model, this demonstrates a timely understanding of the impact of social activities on well-being. Academic research demonstrates that social well-being is closely tied to physical health, a well-known example being the impact of socioeconomic positioning on mortality as demonstrated in the Whitehall Studies, as well as other more recent work by Michael Marmot.[2 3] Though this common understanding has not fully translated into clinical practice and public health. Particularly in the context of publicly funded medical systems like the UK's National Health Service (NHS), resource limitations and unclear evidence on the causal mechanisms between social activities and improved health make it challenging to incorporate social well-being in treatment models.[4]

Over the past decade, one proposed method of addressing this linking up of health and care services is referral out of primary care health systems and in to the community.[5 6] This 'emerging model of care' was alluded to in the NHS 5Year Forward View[7] in the context of healthcare needing to move to a partnership rather than discrete episodes

of treatment. More substantially, social prescribing was recommended as a key resource for primary care, noting that 'non-medical interventions such as social prescribing can contribute to primary care teams meeting the physical, psychological and social care needs of an individual in the round'[8] (p7). Sometimes with alternative descriptors such as 'community referral', 'community links', and 'arts on prescription', these programmes link healthcare to opportunities and events provided by third-sector organisations. A rapid evidence review by the University of York defined '(social) prescribing (as) a way of linking up patients in primary care with sources of support in the community'; however, the authors highlight that there is no agreed definition.[9] Kimberlee[10] suggests that social prescribing consists of a range of different services, from more traditional smoking cessation programmes, and describes social prescribing as 'a route to reducing social exclusion, both for disadvantaged, isolated and vulnerable populations in general, and for people with enduring mental health problems'. (p105).

Although social prescribing is a commonly used term, we use 'social referral' to be as inclusive as possible in describing links between healthcare and third-sector organisations. In cases where a study specifically uses terms like arts on prescription or 'social prescribing', we refer to it as such. We also do not specify primary care as the only source of social referral; we include referrals by other healthcare workers.

Evidence for the effectiveness of social referral services has been characterised as inconclusive.[9] Although there is significant, if piecemeal, investment in social referral programmes, many advocates of their value[7 10] who attempt to summarise the current evidence, and thus address these criticisms, have similarly been inconclusive in evidencing the health, social, or service-related benefits of social referral.[11–15] Mossabir *et al*[13] conducted a scoping review of seven studies on social prescribing and found that although potentially beneficial for psychosocial health, there had been too few empirical studies to draw clear conclusions. The University of York Centre for Reviews and Dissemination[9] goes as far as to argue 'there is little in the way of supporting evidence of effect to inform the commissioning of a social prescribing programme' (p4).

The first step in evaluating any programme is determining what it aims 'to do' and deciding on the measures that will be used to ascertain effectiveness. There has thus far been little reflection on the intended aims of social referral and the measures used to judge whether the aims have been met. Accordingly, our purpose is to summarise the aims and measures of social referral through a review of the literature. Our first objective is to identify the aims of social referral initiatives. Our second objective is to identify the measures used to evaluate whether the aims of social referral were met. This creates a foundation to inform further programme development and evaluation and for theorising the various mechanisms that may, in specified contexts, be responsible for changes in particular outcomes. We can thus better understand what is meant by 'social prescription' with a view to informing evaluations to consider the contexts in which social referral works, for whom and through which mechanisms.[16]

LITERATURE SEARCH METHODOLOGY

As part of the 'Collaborating to Deliver Social Prescribing in Bath and North East Somerset' project, we conducted a review of empirical and grey literature related to 'social prescribing'. We identified PubMed suggested terms associated with social prescribing, as this is the most commonly used term to identify these kinds of community-linking programmes. The final terms were 'social prescribing', 'social prescribing services', 'social prescription', 'social prescriptions', 'community referrals', 'community referred', 'community-referred patients', 'community refers' or 'community-referring physicians'. We used exactly these terms to search each of the following databases: Scopus, Web of Science, PubMed, National Institute for Health and Care Institute (NICE) Evidence Guidelines database and PsycNET for academic peer-reviewed articles. See online supplementary file 1 for a full example search strategy. The term social referral was not included as we defined this term post hoc, to subsume programmes that did not label themselves as social prescribing as well as those that did. Finally, we examined the first five pages of results identified by internet search engine Google to identify grey literature reports related to social prescribing. After the online database search, academic and non-academic literature reference lists were handsearched. Only the academic literature's citations were searched as several of the non-academic reports were not held on an academic database; therefore, citation searches could not be conducted. The initial search, including citations and reference searching, took place in February 2016 and an updated search was conducted in November 2016 to include recent articles and reports. There were no date restrictions applied in either of these searches.

Identified articles were deemed relevant for inclusion if they reported the assessment of a referral programme of patients from a health context to a social context. A health context was considered any form of health or mental care, for example, emergency departments, primary care, and mental health professionals. A social context was considered any form of community programme including cultural programmes, arts classes, or community groups. This excluded programmes evaluating a single programme, for example, a diabetes health management course. We excluded these 'single intervention' studies as by definition social referral programmes are premised on referring an individual to a range of interventions. After searching using these broad criteria, additional inclusion criteria were added due to the unexpected range of study methodologies, including many interview studies focused on clinical or provider perspectives. These criteria included the use of empirical methodology (qualitative, mixed methods or quantitative),

assessment of a patient sample, and the production of a final article or report. This therefore excluded empirical articles that were evaluating the service provider's views of a social referral programme. Reports or articles that were not in their final version (eg, commissioner or funding interim reports) were excluded as were conference reports and book chapters. No language or region restrictions were applied. After identification of relevant articles and reports, we extracted the study type, stated aim(s) and measures of each social referral programme. We categorised each study's aim(s) as mental, health, social, service use, service cost, and/or other and also extracted number of aims and whether a study aimed to address both individual-level and system-level aims. We did not assess study quality as we were not concerned with the results of social referral only the stated aims and measures. We also extracted the social referral programme name, study design, referral criteria, programme location, programme type, number of programme participants, and number of study participants.

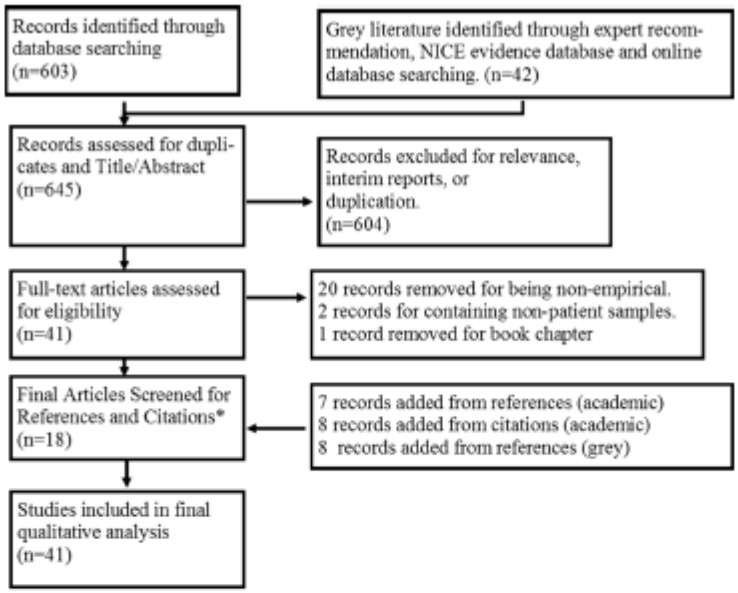
ESR screened all initial articles for title and abstract relevancy, and ENW then read these articles, identified by ESR, for verification that they met inclusion criteria. The first coder, ESR, developed the coding framework and the second coder, ENW, separately coded all articles to this framework. Any differences between the coding of aims or measures, or the inclusion of articles, were subsequently discussed and agreed on. Due to the qualitative nature of the review, we did not calculate percentage agreement.

RESULTS

The initial database search resulted in 645 articles or reports. After duplicate removal, title and abstracts were reviewed according to inclusion and exclusion criteria, 41 articles were identified. On assessment of these full-text articles, 20 were removed for being non-empirical (eg, discussion or review articles that did not evaluate a specific social referral programme but rather provided a general discussion on social referral), two were removed for containing non-patient samples and one was removed as it was a book chapter. After a forwards and backwards citation search, a further 23 articles were identified as relevant. At the initial February 2016 search, six review articles or articles with non-patient samples were also handsearched for references and citations. Three non-academic articles referenced in grey literature reports that may have been relevant could not be found as copies of these reports were not held online, were not available through interlibrary loans and were not held at the British Library. Furthermore after contacting the citing author and place of publication, these articles could still not be found. In total, 41 texts were analysed. See figure 1 for a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of the search strategy and results.

Of the 41 empirical studies, seven were qualitative, 16 were quantitative and 18 employed mixed methodologies. Figure 2 outlines the process of social referral programmes described in these studies. The broad nature of the search led to a broad range of programmes but all followed the basic outline seen in figure 2. There was considerable variation in indicators of need, referral process, and types of activities undertaken. For example, emergency case management as described by Lee and Davenport[17] specifies the population as those who have three or more emergency department visits per month, as well as a list of specific health concerns. Their referral process is nurse-led case management, where they refer to community services as well as other health services. The activities varied including both community as well as more traditional health referrals. In contrast, Stickley and Hui[18] describe a prescriptive arts programme. They do not specify a population, only the referral mechanism. The referral was from a primary or secondary mental health worker. The activity was a 10-week arts programme and the anticipated outcome was personal health improvement. Online supplementary appendix 1 outlines the various types of programmes and study designs. Of the 41 studies, there were 38 unique social referral projects. There were two repeated programmes (Arts on Prescription and the BRIGHT trial); however, the four studies were all individual evaluations of these services. As well, the Health Trainer and Social Prescribing Service[19] was based on a previous pilot of the CHAT programme.[12] The majority of these texts described either a social prescription programme or an emergency department case management programme. All of the social prescribing programmes were set in the UK. The emergency department case management programmes were located in the USA, UK, Canada, and Taiwan. All studies included only adult populations with study size ranging from 4 to 784. Patient samples varied greatly, from kidney patients to elderly adults. Programme size also greatly varied from 12 to 1848 referrals. See online supplementary appendix 1 and 2 for more details.

Table 1 outlines the aims of the programmes described in the empirical studies. The stated aims were those listed in the individual studies, while the core aims were derived by grouping together similar aims across programmes. The core aims were then grouped in relation to the level at which the intervention was aimed: individual or system. The core individual aims identified included improved mental well-being, improved physical well-being, and improved social well-being. The core system-level aims included optimised health service use and decreased health service cost. Only nine studies stated a single aim. The majority of studies thus stated multiple aims: 16 stated two, 10 stated three, four stated four and one study stated five aims. Nineteen studies focused on both individual-level and system-level outcomes (see online supplementary appendix 2 for full details). Improved mental well-being was the most common core aim, with 25 of 41 studies. Physical well-being, social well-being, and optimised service use were also frequently cited with 16, 21 and 23 studies, respectively. Six studies addressed the least common core aim of cost savings.



*Additional articles (e.g. review and non-empirical papers) that did not meet inclusion criteria in previous search stages were also hand-searched for citations and references.

Figure 1 The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for the literature search strategy for social referral programmes. The main criterion for inclusion was an empirical assessment of a programme that contained a patient referral out of the healthcare system and into the community or voluntary system. Six hundred and forty-five articles and reports were initially identified and assessed for duplication and relevance. Forty-one articles and reports were then assessed for full-text eligibility. Eighteen articles or reports were identified. The citations and reference lists for the academic articles were searched for additional literature, alongside other non-eligible review papers, as well as the reference lists of the non-academic reports. This resulted in 23 articles further identified as relevant. Finally, 41 studies were included in the qualitative synthesis. NICE refers to the National Institue for Health and Care Excellence.

The mental well-being core aim was generally characterised by mental health or general well-being. Improved psychosocial state was considered to be both related to social and mental well-being. Physical well-being included both general health and the improvement of long-term health conditions, like kidney disease. Social well-being included improvements in social and community engagement and quality of life. Health

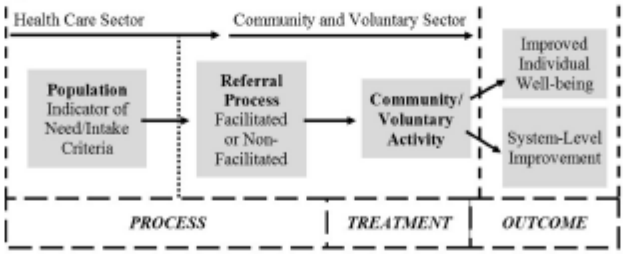


Figure 2 A summary of the social referral process identified in the literature search. All programmes' participants were identified by various indicators of need, for example, low-level mental health conditions within the healthcare sector. The participants were then provided with either a facilitated or non-facilitated referral to a community or voluntary activity. Patient identification and referral represent the 'process' while the activity represents the 'treatment' of social referral programmes. Finally, the proposed outcomes included either improved individual well-being, for example, mental well-being, and/or system-level improvement, for example, reallocated healthcare resources.

Table 1 Summary of aims of social referral programmes (n=41)

Aim level	Core aim	Stated aim	Number of references
Individual-level aim	Improved mental well-being	To enhance skills/behaviours that improve mental well-being.[29]	
		To help individuals retain/recover functional capacity to study or work.[30]	
		To improve/address psychosocial health.[31–35]	25
		To improve mental health and well-being.[5 16 20 23 29 36–46]	
		To improve patient quality of life.[46 47]	
		To improve resilience, confidence, and self-esteem.[44 46]	
		To improve spiritual well-being.[5]	
		To support emotional needs.[49]	
	Improved physical well-being	To empower and support individuals to choose a healthier lifestyle.[46]	
		To improve physical health and well-being.[5 17 20 23 31 37 38 40 42 50–53]	16
		To improve self-assessed health status.[54]	
		To support the self-management of long-term health conditions.[38 50 55]	
	Improved social well-being	To increase connection to community-based support.[29 37]	21
		To improve/address psychosocial health.[31–35]	
		To improve resilience, confidence, and self-esteem.[46]	
		To improve social inclusion/engagement.[20 23 30 32 36 41]	
		To improve social well-being[40 42 52]	
		To support social needs/outcomes.[19 36 49 53 56]	
	Other	To address practical needs, for example, employment.[49]	2
		To improve connection to nature.[23]	
System-level aim	Optimised health service use	To broaden health service provision in the community.[12]	23
		To improve service use.[32]	
		To increase take-up of community activities.[29 36 44]	
		To optimise healthcare coordination.[57]	
		To provide appropriate arts course recommendations.[44]	
		To provide better management of psychosocial problems in primary care.[47]	
		To reduce emergency department use/acute hospital care.[17 35 37 51 58 59]	
		To reduce health service use.[21 39 42 53 54 57]	
		To reduce hospital care use.[22 36 59]	
		To reduce primary care service use.[16 34 37 38]	
		To support the self-management of long-term physical or mental health conditions.[44 50 55]	
	Decreased health service cost	To reduce cost associated with long-term health conditions.[50]	6
		To reduce health services costs.[5 21 35 42 53]	
	Other	To reduce environmental cost (carbon footprint).[21]	1

Aims of social referral programmes, not study aims.

service use and cost aims included reductions in emergency department use, general practitioner (GP) use, hospital stay length and other forms of primary care costs. The service use aim also included instances where researchers were aiming to increase the uptake of community services. See online supplementary appendix 2 for more detail on aims.

Table 2 outlines the measures and methods used to evaluate the social referral projects by frequency. Across all aims, these included administrative data/analysis, physical health questionnaires, mental health diagnostic measures, qualitative assessments and social/behavioural questionnaires. Across the 41 studies and reports, 154 different kinds of measures or methods of evaluation were identified (see online supplementary appendix 2). Twenty-one measures or methods were used more than once; however, many of these were forms of administrative data counts. The most commonly used scale was the Warwick-Edinburgh Mental Well-being Scale, used in nine studies.

Table 2 Measures and methods used in studies/reports of social referral by frequency (n=41)

Measure/method	No of studies/reports using measure/method	Examples of programme aims addressed*
Semistructured interviews to explore patient experience	14	NA†
Warwick Edinburgh Mental Well-being Scale (14 or 7 item)	9	Improved mental well-being Improved physical well-being Improved social well-being
Number of GP appointments (administrative)	6	Optimised health service use Reduced health service cost Improved physical well-being
Short case description of participant experience	6	Improved physical well-being Improved social well-being Optimised health service use
Emergency department admissions/Hospital Episode Statistics (administrative)	6	Optimised health service use
Demographic questions	5	Improved mental well-being
Cost analysis	5	Reduced health service cost Optimised health service use
Hospital Anxiety and Depression Scale	5	Improved mental well-being Improved physical well-being
Focus group with patients to explore patient outcomes	4	NA‡
General Health Questionnaire-12	3	Improved mental well-being Improved physical well-being
No. of secondary referrals (administrative)	3	Optimised health service use Reduced health service cost
Geriatric Depression Scale	2	Improved mental well-being
Focus group with family members who engaged with the service to explore service experience	2	NA‡
Hospital admissions length (administrative)	2	Optimised health service use
Reason for referral	2	Improved mental well-being Optimised health service use
Referral records (eg, what activities were referred to)	2	Improved social well-being
Social Return on Investment Analysis	2	Reduced health service cost Improved mental well-being
Work and Social Adjustment Scale	2	Improved social well-being
No. of Hospital Admissions (administrative)	2	Optimised health service use
No. of prescriptions for psychosocial reasons (administrative)	2	Optimised health service use Improved mental well-being

Where the measure or method was used in n>1 report or study.
*These are only example aims because it was not always clear how each aim and measure matched up.
†Not applicable as the qualitative semistructured interviews and focus groups were exploratory and did not have a specific programme aim to measure.

DISCUSSION

Examination of the aims of studies seeking to evaluate social referral initiatives and the measures used to evaluate their outcome has revealed extensive heterogeneity. This is unsurprising considering the variability in populations and types of programmes and is not problematic per se. We will discuss the various aims of social referral and the implications of the variety of measures used before considering what this variability means for the future of social referral programmes. In doing so, it is important to reiterate the hugely varied nature of the events and opportunities to which people are being referred, as well as the substantial variety of recipients of this referral. While we expect variation in programme aims and measures, these varied programmes were included because they all aimed to link individuals with community and healthcare services. It is therefore reasonable to assume that there would be some kind of consistency in the measures used to address particular aims.

Aims of social referral

The vast majority of studies, 32 out of the total 41, included multiple aims. Nineteen of these were concerned with

both individual-level and system-level outcomes (see table 1 and online supplementary appendix 2), for example, mental well-being and health service costs. While a single study containing aims at individual and system levels is not problematic as such, what is problematic is the lack of articulation of the presumed causal pathways from the treatment programme to improved individual health and to better healthcare resource allocation. As a thought experiment, an individual who is a frequent health service user and has poor control over their diabetic care could, in theory, be empowered by a social referral service and continue high levels of primary care access as they take greater ownership of their health. Indeed a few studies have found an uptake in medical service use post-social referral.[20–22] It is also important to note that when reviewing the grey literature, and indeed some of the academic literature as well, the aims of the programme were not always clearly stated. It is reasonable for programmes to try to address multiple aims; however, it is not acceptable for these programmes not to theorise, test, and critically evaluate the relationship between them.

Measures of social referral

Measuring what 'works' is inherently linked to defining what these programmes intend to do and requires meaningful, specific, and comparable indices. The diversity of measures evident in social referral initiatives, often associated with a series of vaguely similar aims, suggests that what programmes are aiming to do is often different despite having notionally similar programme structures. Additionally of course it is important to take into account the role of population type and activity type in how aims are translated in to measures. However, as seen in table 2, measures used in social referral initiatives are considerably more plentiful than their aims. For example, Bragg *et al*[23] used 12 different tools in their evaluation of an ecotherapy programme. The multiple measures both within and between studies render comparability between studies, even those addressing the same or similar aims, impossible. Similarly, we could not meaningfully narrow them to provide recommendations on preferred measures. Where there were multiple aims, papers rarely stated which measure was meant to address which aim. While we might infer that administrative counts of GP visits would measure GP use, the assumed relationship between number of GP visits and physical well-being is less clear. Clarity of reporting in the hypothesised relationship between aims and outcome measures is vital in understanding the causal mechanisms that link a programme with its outcomes. From one perspective, measuring the same outcome in several ways could lead to a more robust proof of effect. In theory, this could lead to a stronger evidence base about the effect of social referral on individual-level and system-level outcomes. A less generous explanation behind the proliferation of measures is that researchers and evaluators do not have a definitive understanding of how exactly the aim of their social referral service can translate in to measures. Where the aims are not clearly set out, it may be that they are not being communicated well but the possible explanation that the aims are unknown or unclear cannot be ruled out. It certainly suggests that one of the essential building blocks for an evaluation of a complex health system,[24] that is, establishing the current evidence base, has not been undertaken and/or understood. Establishing the evidence base constitutes a crucial springboard for developing hypotheses as to the mechanisms through which social prescribing programmes might improve social well-being and, ultimately, physical and health outcomes. Identification with the group, for example, rather than simply engaging in group activities may be one such mechanism.[25]

In the final analysis, while there is a notable policy push for the implementation of social referral programmes, definitive and systematic evaluations of social referral programmes are not possible while aims and measures are so inconsistent. As a caveat, one can expect that where populations and activities vary one can expect different measures. However, where social referral programmes aim to do similar things, measures that are similar should follow, for example, the Short Warwick-Edinburgh Mental Well-being Scale is not population, nor activity specific. We hope that this review provides a first step towards categorising the aims of social referral programmes, that is, to improve physical, mental and social health, as well as reducing costs and improving healthcare resource allocation. Although these aims are broad, they provide a framework for highlighting what these programmes intend to do, and not do, and identifying which measures might best be used to assess different types of aims. This would be a start in applying a more consistent methodology.

The solution to the issue of aim and measurement variability in programmes is not to give up on social referral in general. Certainly, the incorporation of social and mental well-being within traditional biomedical health systems seems an essential step in tackling relatively recent problems in healthcare, for example, services for ageing populations, and may create new opportunities for people who are stagnated in their ability to access services that improve their health. However, at this time, despite policy claims of value and claims of the effectiveness of individual programmes, reviews of these programmes are clear that we do not have evidence that this is the case.[9 12–15 26–28] We would argue that while aims and measures remain diffuse and the links between them undertheorised and underspecified that we actually *cannot* know that this is the case. We call on researchers and evaluators alike to consider the active ingredients of their programmes and in doing so echo a similar call made by the University of York asking, simply, for whom, in what context, how, and why do they intend to prescribe social activities[9]? And while these can be challenging to answer, if we do not know the answers to these simple questions, how can we possibly prepare a prescription?

Strengths and weaknesses

Although this review has been systematically conducted providing a transparent account of the process, we cannot guarantee this has included all relevant social referral programmes. Social prescribing is a generalised UK region-specific term for medical-based referral to non-medical services. There are likely social referral-like programmes in other countries that are not easily identified. Every effort was made to be as inclusive as possible in phrasing but there will inevitably be some studies missed. Conversely, the strength of our analysis is our inclusion of both grey and academic literature. By including non-academic reports, we analysed valuable literature that would normally not be included in reviews. As well, this review is a first step in creating consistency and justification for the inclusion of social referral programmes in broader nationwide initiatives to address the social ills of health. The contribution of our approach to reviewing social referral is valuable due to its focus on aims and measures rather than, as is the case in other reviews, the outcomes of programmes.

CONCLUSION

This review aimed to analyse and summarise the aims and measures used in the evaluation of social referral programmes. Social referral is variously described as social prescribing, community referral, and emergency case management among other terms. We found great variation in the aims of these projects including aims to improve mental well-being, physical health, social well-being, and costs savings. We further found that measures used to analyse these aims were highly varied. We would suggest that a next step to addressing the social determinants of health in primary and secondary care is to derive more differentiated and concrete definitions of social referral that more specifically reflect what practitioners and commissioners intend for programmes to achieve and thus to dispense with a general notion of social referral often uncritically considered as the 'golden child' of cost savings and improved mental health. However, by setting clear aims and using appropriate measures, social referral can move beyond pilot studies and in to general practice. To that end, we must endeavour to respond to Walt Disney's call to 'diagnose and prepare the prescription'.[1]

Acknowledgements The authors would like to acknowledge the support of the 'Collaborating to Deliver Social Prescribing in Bath and North East Somerset' Project Team, in particular Developing Health & Independence, the Wellbeing College, Second Step, Bath & North East Somerset Council, led by Digital Algorithms Ltd.

Contributors ESR, JB and HD designed the study protocol. ESR conducted the database searching, while ESR and ENW conducted the data extraction. The report was written by ESR and JB. All authors edited the manuscript.

Funding This work was supported by Innovate UK, project code: 102412-399209.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

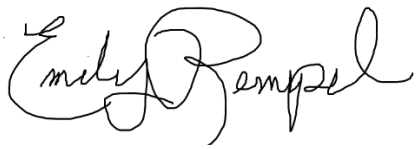
Data sharing statement Full coding guidelines and summaries for all articles included can be found in the supplementary appendices 1 and 2.

REFERENCES

1. Disney W. Film entertainment and community life. *J Am Med Assoc* 1958;167:1342–5.
2. Marmot M. *UCL Institute of Health Equity. Review of social determinants and the health divide in the WHO European Region: final report*. Copenhagen: WHO Regional Office for Europe, 2014.
3. Marmot MG, Stansfeld S, Patel C, *et al*. Health inequalities among British civil servants: the Whitehall II study. *The Lancet* 1991;337:1387–93.
4. Marshall M. A precious jewel—the role of general practice in the English NHS. *N Engl J Med* 2015;372:893–7.
5. Kimberlee R, Ward R, Jones M, *et al*. *Proving our value: Measuring the economic impact of Wellspring Health Living Centre's social prescribing Wellbeing Programme for low level mental health issues encountered by GP services*. Bristol: University of the West of England, 2014.
6. Kimberlee R. *Developing a Social Prescribing approach for Bristol*. University of the West of England, 2013.
7. NHS. *NHS Five Year Forward View*. England: National Health Institute, 2014.
8. Dyson B. *Improving general practice: A call to action – phase one report*. London, UK: NHS England, 2014.
9. The University of York Centre for Reviews and Dissemination. *Evidence to inform the commissioning of social prescribing*. York: University of York, 2015.
10. Kimberlee R. What is social prescribing? *ASSRJ* 2015;2:102–10.
11. Bungay H, Clift S. Arts on prescription: a review of practice in the U.K. *Perspect Public Health* 2010;130:277–81.
12. South J, Higgins TJ, Woodall J, *et al*. Can social prescribing provide the missing link? *Prim Health Care Res Dev* 2008;9:310.
13. Mossabir R, Morris R, Kennedy A, *et al*. A scoping review to understand the effectiveness of linking schemes from healthcare providers to community resources to improve the health and well-being of people with long-term conditions. *Health Soc Care Community* 2015;23:467–84.
14. Kumar GS, Klein R. Effectiveness of case management strategies in reducing emergency department visits in frequent user patient populations: a systematic review. *J Emerg Med* 2013;44:717–29.
15. Soril LJ, Leggett LE, Lorenzetti DL, *et al*. Reducing frequent visits to the emergency department: a systematic review of interventions. *PLoS One* 2015;10:e0123660.
16. Pawson R, Tilley N. *Realistic evaluation*. London: SAGE 1997.
17. Lee KH, Davenport L. Can case management interventions reduce the number of emergency department visits by frequent users? *Health Care Manag* 2006;25:155–9.
18. Stickley T, Hui A. Social prescribing through arts on prescription in a U.K. city: participants' perspectives (part 1). *Public Health* 2012;126:574–9.
19. White J, Kinsella K, South J. *An evaluation of social prescribing health trainers in South and West Bradford*. Leeds: Leeds Metropolitan University, 2010.
20. Cohen GD, Perlstein S, Chapline J, *et al*. The impact of professionally conducted cultural programs on the physical health, mental health, and social functioning of older adults. *Gerontologist* 2006;46:726–34.
21. Maughan DL, Patel A, Parveen T, *et al*. Primary-care-based social prescribing for mental health: an analysis of financial and environmental sustainability. *Prim Health Care Res Dev* 2016;17:114–21.
22. Gupta K, Coupland L, Fottrell E. A two-year review of an 'open access' multidisciplinary community psychiatric service for the elderly. *J Geriatr Psychiatry* 1996;11:795–9.
23. Bragg R, Wood C, Barton J. *Ecominds effects on mental wellbeing: an evaluation for mind*. Essex: University of Essex, 2013.
24. Lamont T, Barber N, de Pury J, *et al*. New approaches to evaluating complex health and care systems. *BMJ* 2016;352:1–5.

25. Sani F, Madhok V, Norbury M, *et al*. Greater number of group identifications is associated with healthier behaviour: Evidence from a Scottish community sample. *Br J Health Psychol* 2015;20:466–81.
26. Thomson LJ, Camic PM, Chatterjee HJ. *Social prescribing: A review of community referral schemes*. London: University College London, 2015.
27. Kinsella S. *Social prescribing: A review of the evidence*. Wirral: Wirral Council Business & Public Health Intelligence Team, 2015.
28. Bickerdike L, Booth A, Wilson PM, *et al*. Social prescribing: less rhetoric and more reality. A systematic review of the evidence. *BMJ Open* 2017;7:e013384.
29. Friedli L, Themessl-Huber M, Butchart M. *Evaluation of Dundee Equally Well Sources of Support: Social prescribing in Maryfield Evaluation report four*. Dundee: Equally Well Dundee, 2012.
30. Garety PA, Craig TK, Dunn G, *et al*. Specialised care for early psychosis: symptoms, social functioning and patient satisfaction: randomised controlled trial. *Br J Psychiatry* 2006;188:37–45.
31. Greaves CJ, Farbus L. Effects of creative and social activity on the health and well-being of socially isolated older people: outcomes from a multi-method observational study. *J R Soc Promot Health* 2006;126:134–42.
32. Crawford M, Rutter D, Price K, *et al*. *Learning the lessons: a multi-method evaluation of dedicated community-based services for people with personality disorder: National Co-ordinating Centre for NHS Service Delivery and Organisation R&D (NCCSDO)*, 2007.
33. Faulkner M. Supporting the psychosocial needs of patients in general practice: the role of a voluntary referral service. *Patient Educ Couns* 2004;52:41–6.
34. Grayer J, Cape J, Orpwood L, *et al*. Facilitating access to voluntary and community services for patients with psychosocial problems: a before-after evaluation. *BMC Fam Pract* 2008;9:27.
35. Okin RL, Boccellari A, Azocar F, *et al*. The effects of clinical case management on hospital service use among ED frequent users. *Am J Emerg Med* 2000;18:603–8.
36. Stickley T, Eades M. Arts on prescription: a qualitative outcomes study. *Public Health* 2013;127:727–34.
37. Innovation Unit. *Wigan community link worker service evaluation*. London: Innovation Unit, 2016.
38. City and Hackney Clinical Commissioning Group, University of East London. *Shine 2014 final report social prescribing: Integrating GP and community assets for health*. London: University of East London, 2014.
39. Newcastle West Clinical Commissioning Group. *Social prescribing for mental health: an integrated approach project report*. Newcastle: Newcastle West Clinical Commissioning Group, 2014.
40. Jones M, Kimberlee R, Deave T, *et al*. The role of community centre-based arts, leisure and social activities in promoting adult well-being and healthy lifestyles. *Int J Environ Res Public Health* 2013;10:1948–62.
41. Baker K, Irving A. Co-producing approaches to the management of Dementia through social prescribing. *Soc Policy Adm* 2016;50:379–97.
42. ERS Research and Consultancy. *Newcastle social prescribing project: final report*. Newcastle upon Tyne: ERS Research and Consultancy, 2013.
43. Morton L, Ferguson M, Baty F. Improving wellbeing and self-efficacy by social prescription. *Public Health* 2015;129:286–9.
44. Huxley P. *Arts on prescription: an evaluation*. UK 1997.
45. The Care Forum. *New routes: end of project report*. Bristol, UK: The Care Forum, 2015.
46. Kilroy A, Garner C, Parkinson C, *et al*. *Towards Transformation: exploring the impact of culture, creativity and the arts on health and wellbeing: a consultation report for the critical friends event*. Manchester, UK: Arts for Health, Manchester Metropolitan University, 2007.
47. Grant C, Goodenough T, Harvey I, *et al*. A randomised controlled trial and economic evaluation of a referrals facilitator between primary care and the voluntary sector. *BMJ* 2000;320:419–23.
48. White M, Salamon E. *An interim evaluation of the 'Arts For Well-being' social prescribing scheme in County Durham*. Durham: Centre for Medical Humanities, Durham University, 2010.
49. Ramsbottom H, Volpe J, Pantelic V. *Social prescribing: a model for partnership working between primary care and the voluntary sector*. London: Age UK, n.d.
50. Blakeman T, Blickem C, Kennedy A, *et al*. Effect of information and telephone-guided access to community support for people with chronic kidney disease: randomised controlled trial. *PLoS One* 2014;9:e109135.
51. Liao M-C, Chen L-K, Chou M-Y, *et al*. Effectiveness of Comprehensive Geriatric assessment-based intervention to reduce Frequent Emergency Department visits: a report of four cases. *Int J Gerontol* 2012;6:131–3.
52. Vogelpoel N, Jarrold K. Social prescription and the role of participatory arts programmes for older people with sensory impairments. *J Integr Care* 2014;22:39–50.
53. Dayson C, Bashir N. *The social and economic impact of the Rotherham Social Prescribing Pilot: Summary evaluation report*. Sheffield: Sheffield Hallam University, 2014.
54. Reinius P, Johansson M, Fjellner A, *et al*. A telephone-based case-management intervention reduces healthcare utilization for frequent emergency department visitors. *Eur J Emerg Med* 2013;20:327–34.
55. Blickem C, Kennedy A, Jariwala P, *et al*. Aligning everyday life priorities with people's self-management support networks: An exploration of the work and implementation of a needs-led telephone-based support systen. *BMC Health Services Res* 2014;14:1–12.
56. Goodhart C, Layzell S, Cook A, *et al*. Family support in general practice. *J R Soc Med* 1999;92:525–8.
57. Hudon C, Chouinard MC, Diadiou F, *et al*. Case management in primary care for frequent users of health care services with chronic diseases: a Qualitative Study of Patient and Family Experience. *Ann Fam Med* 2015;13:523–8.
58. Skinner J, Carter L, Haxton C. Case management of patients who frequently present to a Scottish emergency department. *Emerg Med J* 2009;26:103–5.
59. Tadros AS, Castillo EM, Chan TC, *et al*. Effects of an emergency medical services-based resource access program on frequent users of health services. *Prehosp Emerg Care* 2012;16:541–7.

ARTICLE: REALISING THE BENEFITS OF INTEGRATED DATA FOR LOCAL POLICY-MAKING

This declaration concerns the article entitled:									
Realising the benefits of integrated data for local policy-making									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**	✓	**Accepted**		**Published**	
Publication details (reference)	Durrant H., Barnett J., and Rempel E.S. (2018). Realising the benefits of integrated data for local policy-making**.** *Politics and Governance, (under review).*								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to the methodology design, work, and presentation of data. Formulation of ideas: HD and JB (100%) primarily formulated the idea of the research. Design of methodology: ESR helped design two of the four case studies (30%) with HD and JB conducting the remaining 70% of the methodology design. Experimental work: ESR helped conduct two of the four case studies mentioned in the methodology (10%) with HD and JB identifying the case study themes and conducting the other two case studies (90%). Presentation of data in journal format: HD wrote up the results of the research (70%) alongside JB (20%) with editing and writing provided by ESR (10%).								
Statement from Candidate	This paper reports on original research I collaborated on during the period of my Higher Degree by Research candidature.								
Signed						**Date**	01/06/2018		

ABSTRACT

This paper presents findings from case studies of local government activity to realise the benefits of big data for policy. Through participatory action research with two local statutory authorities in the South West of England, we explored the experience of identifying, integrating and analysing multiple and diverse forms of data to generate insights on live policy priorities and inform decision-making. We reveal the significance of both data production and policy-making contexts in explaining how big data can be called upon and enacted in policy processes.

1. INTRODUCTION

The claims made for big data in business contexts are well established (e.g. Mayer-Schönberger & Cukier, 2013). More recently, attention has turned to the potential of big data for policy-making settings (e.g. Parliament.UK, 2015), and the challenges involved in harnessing this potential to realise policy aims and objectives for the public good (Schintler & Kulkarni, 2014; Kennedy, Moss, Birchall, & Moshonas, 2015; Janssen, Konopnicki, Snowdon & Adegboyega, 2017). Questions have been raised about how and where in the 'policy cycle' big data-derived analysis should feed in (Höchtl, Parycek & Schöllhammer, 2016), with increasing emphasis being placed on the role that data can play in predicting need and defining policy priorities for the future (Malomo & Sena, 2016; Giest, 2017). This work usefully disaggregates the applications of data, moves beyond rhetoric and opens up thinking about the spaces for data science to inform policy-making.

However, policy-making process are not straightforward or linear, and there is a need to theorise the social contexts of both data production and policy-making to understand the boundaries and barriers to big data for policy in practice. We reveal the temporally-specific and contingent ways in which data are articulated in the demand for evidence, and discuss how the practices and preoccupations of policy-making both shape and are being shaped by the promise of data.

We begin (in section two) by rehearsing the claims that have been made about big data, and that have sought to give this ubiquitous but simultaneously elusive term some definitional clarity. We focus on claims made about the promise of data for policy-making, and problematize assumptions of linear and rational policy-making processes into and through which data science can flow. We propose a counter theory of policy-making as struggles over the right to advance ideas about policy; why it is needed, what it should do, for whom, how and to what end (Carmel & Papadopoulos, 2003). We argue that it is in these deeply value-laden and political contexts that data are produced and repurposed, and insights are allowed, or otherwise, to be admitted as a form of evidence.

Section 3 briefly describes the participatory action research approach adopted in this project and details the partnership and processes by which the project progressed. Section four presents findings and reflections from the project; focusing on the ways in which data is constituted as relevant to policy-making, the terms on which its use is resisted; and the importance of relationships of trust to underpin data processes in practice. We conclude in section five by discussing the significance of the social context of both data generation and policy-making to explain what can actually be done with data in policy settings.

2. BIG DATA and POLICY MAKING

Typically, when big data is spoken about, and definitions emerge, two interrelated phenomena are invoked; changes to the nature of data and the changing approach to data analytics (boyd & Crawford, 2012). This dual focus is seen in the definition of Schroeder (2014, p. 165), "Big data can be defined as research that is made possible by means of the capture, aggregation, and manipulation of data about a given phenomenon on an unprecedented scale and scope".

Attempts to define big data have tended to focused on its characteristics; including but not limited to its volume, variety and velocity (see Malomo & Sena, 2016). Consideration of the sources of data is also useful in grounding our understanding of what is commonly considered to constitute big data. Data are being generated

from a greater variety of sources than ever before (Kitchin, 2014). Some of this data is what Mayer-Schönberger & Cukier (2013, p. :113) refer to as "data exhaust"; the by-product of peoples digital activities and interactions, repurposed to another end. Large administrative data sets, collected for one purpose and used and connected for others, are a further example of what can be considered as big data (Malomo & Sena, 2016; Janssen et al, 2017).

The identification of a wide array of characteristics and sources of big data conveys a sense of its ubiquity, but also the extent to which it defies definitional clarity. Indeed, maybe 'big data' is not a thing, but, as Markus and Topi (2015, p. :3) contend "a cluster or assemblage of data-related ideas, resources and practices". Such a definition allows for a consideration of the human actions and interactions involved in the identification, interrogation, interpretation and application of data (ibid, p. : 9). This paper is concerned with what these human actions might consist of in the context of big data for policy-making.

Governments, at all levels, have unique access to diverse data generated on the people and communities they serve, and there is a growing literature on the application of data in policy settings (Malomo & Sena, 2016; Janssen et al, 2017). Promise and potential are the watchwords and although short comings and dangers are highlighted, the claims made are often expansive. Stephan Shakespeare, in his review of Public Sector Information, enthusiastically asserts that "…from data we will get the cure for cancer as well as better hospitals; schools that adapt to children's needs making them happier and smarter; better policing and safer homes; and of course jobs." (Shakespeare, 2013, p. :5).

In 2015 the UK Parliament identified harnessing the benefits of big data as a key issue for governments; describing data as "the new oil" (www.parliament.uk, 2015). However, what is less clear is by what processes and practices it can meaningfully grease the wheels of decision-making in policy settings. Recent scholarship has sought to identify opportunities for big data insights to inform policy-making, by focusing on the stages of the policy cycle most amenable to

injections of data-derived evidence. Höchtl et al (2016) journey through the steps involved in policy making – e.g. agenda setting and discussion, policy formation and decision-making, implementation, etc. - providing reflections on the potential contribution of big data to each. They particularly highlight the possibility for real time data processing to enable continuous evaluation throughout the process. Giest (2017) explores government use of a range of administrative and real-time data to design and customise policies. She highlights the value of these data to agenda setting and policy implementation. Malomo & Sena (2016) describe a case study of using integrated data in local government and highlight the benefits of big data for predicting need and effectively targeting services.

The studies usefully break down and compartmentalise the different functions of big data for policy making – options appraisal, predictive analysis, real-time evaluation etc. However, they tend to overplay the extent to which policy-making proceeds stepwise, through a series of linear stages, and understate the challenges associated with the straightforward inflow of any kind of information and evidence (Cairney & Hekkila, 2014).

Rather than seeing policy-making as a linear and rational process, we start from the premise that policy-making is the variable outcome of consensus, negotiation, contestation or co-option of ideas about what is to be done, by whom, how and for what purpose (Carmel & Papadopoulos, 2003). Ideas embodied in narratives of causation compete for the right to be accepted, and power and context influences the strength of the narrative to succeed (Stone, 1989; Jessop, 2009). Policy-making is a messy process in which conflicting ideas and policies are brought forward, debated, and implemented but not always in that order.

Scholarship is emerging on how data and data technologies fit into a narrative-conflict view of policy-making. Kettl (2016) emphasises that the nonlinear nature of policy-making problematises the assumption that data is used simply as evidence to make the best policy choice. They argue that good data analysis is useless without a good narrative. In contrast, Janssen & Helbig (2016) argue that

data technologies have great potential to interrupt the status quo and revolutionise policy-making.

In summary then, the enthusiasm within government for realising the potential of big, integrated forms of data has primarily focused on the potential of technical innovations. However, processes of data science in policy settings are embedded in dynamic, multifaceted, and deeply political contexts of problem definition, evidence interpretation, solution identification and decision making. These settings materially affect the ways in which big data is called upon and able to impact decision making. We use case studies of local government activity around integrated data to consider how data informs policy-making processes: how the practices and preoccupations of the policy process define and shape the generation and use of data science; and how integrated data, as one form of evidence generation, shapes and redefines these policy practices.

3. METHODS

This paper presents findings from a co-produced project to realise the benefits of integrated data to inform policy development and practice at the local level. The core project team included three researchers from the University of Bath Institute for Policy Research and senior policy officials from two local statutory bodies within the South West of England. In the course of the project the team engaged with service managers from other departments, and with other policy-making bodies and civil society organisations within the region.

The project was grounded in the principles of participatory action research (Coghlan & Brydon-Miller, 2014; Bergold & Thomas, 2012). Within the terms of a formal collaboration agreement between the institutions, working research practices emphasised equal collaboration between researchers and practitioners, trust and discretion in communication and the production of shared knowledge. The primary aim was to establish, test and evaluate a process to change the culture of data use within and across public services. The team identified mechanisms (technologies and processes) for safely linking anonymised data and sought to

realise the benefits of integrated data by generating new insights into public need and service effectiveness. This approach was applied to four contemporary policy priorities for the statutory authorities. These case studies contributed to the development of policy and practice, and to understanding the processes and techniques that realise the benefits of data for local government.

Table 1 outlines the four case studies and the associated data sources used to inform decision-making.

Table 1: Policy priorities, aims and data sources

Policy Priorities	**Policy Aims**	**Indicative Data Sources**
Financial hardship	To understand the consequences of economic downturn and austerity for financial wellbeing.	*Individual-level:* • Time-series: social benefit claims, employment status, household composition, disability status • Demographic information *Aggregate-level (Lower Super Output Area):* • Debt (County Court Judgements) • Household Composition • Social benefits claims • Tax Credits • Income deprivation
Community health services	Review community health services for patients with a particular chronic condition to understand the efficacy of these services and the effect on health outcomes.	*Individual-level:* • Time-series: secondary care records, in-patient admissions, out-patient appointments, co-morbidities and clinical test results. • Attendance at community care services • Demographic information *Aggregate-level (GP Surgery):* • Patient population • Health checks *Additional data collected:* • Patient illness perceptions and experience of services
Wellbeing services	Review and redesign community wellbeing services.	*Individual-level:* • Wellbeing service administrative records: participant numbers, dates and service location. • Case notes • Evaluation and outcomes. *Additional data collected:* • Interest in wellbeing services

		• Various measures of personal wellbeing • Demographic information • Provider experiences of delivering wellbeing services
Education services	Understand changed profiles of demand and redesign education services.	*Individual-level:* • Time-series: school and academies census • Pupil demographic information • Educational needs and status • Free school meals eligibility

In each case, the project progressed through discussion to understand the policy issues and context; define policy questions of interest; identify and access potential sources of data; and interpret analysis. Insights from the analysis often raised additional questions, and policy questions were refined and additional data sought accordingly. Thus these processes were highly iterative and required researchers and policy officials to adapt to changing configurations of policy context, research questions and data.

4. FINDINGS and REFLECTIONS

In drawing together the case studies and seeking to explore the interactions between the policy context, policy questions and data integration practices, we present findings and reflections under three themes. Firstly, we consider the way in which the relevance of data is constituted in policy settings, as a function of its perceived value in answering policy questions. Secondly, we explore the conditions under which data applications to policy are resisted. Finally, we reveal significant aspects of the relationships between different interested parties where data and policy-making intersect.

4.1 Relevance of data to address policy questions

In using integrated data in local government settings, policy questions, not data, were the starting point for data projects. Whether the issue was financial hardship, designing health and wellbeing services or education service provision, it was the policy questions and context that defined the scope for data to inform decision

making. Data did not "speak for themselves" (Anderson, 2008; see also Mayer-Schonberger & Cukier, 2013). Its potential utility to policy-making was realised where it was admitted (along with other evidence) as a response to a policy enquiry. In other words, the value, or otherwise, of data was constituted only in relation to policy questions, and the weight of data, typically associated with the big data phenomenon, did not unproblematically transfer into a weight of evidence (Schintler & Kulkarni, 2014). Even where there were large volumes of data, the relevance of that data to local policy-making was contingent on the perceived problem and the questions that policy makers could allow about it.

Having said that, the cases do illustrate how a keen interest in the power of data, particularly the potential of combining multiple forms of disparate data, is reinvigorating and reshaping the demand for evidence in policy-making processes at the local level. Policy partners were keen to identify and explore the benefits of the vast amounts of data routinely collected to inform service development, and were, in some cases, open to broadening the options for policy change in light of the subsequent insights.

There was sometimes an absence of data deemed relevant to address particular policy questions. A review of local health services explored patient pathways and outcomes through services relating to a particular condition. In a routine appraisal of these services policy officials were interested in understanding barriers to and enablers of service take-up. They had a clear view about the nature of the policy problem: low levels of service take-up among certain patient groups in particular areas - and a set of questions predicated on assumptions about policy options for service improvement.

However, project discussions with the research team led them to broaden their enquiries. They commissioned a Rapid Evidence Review (RER) to extend their understanding of the factors influencing service uptake. The RER raised explanations for low service take up and variation in service performance that were not previously part of the scope of the data project. This called into question

the sufficiency of the data that had previously been designated as relevant to informing the policy question. Policy makers became aware that data routinely collected and available on these services largely served to facilitate service administration and audit (e.g. by providing information on volume of provision, attendance and dates) rather than understanding reasons for low service take up review performance. They recognised gaps in the data relating to patient experience, as well as patient health management behaviours. In this case they decided to collect additional survey data, which was combined at the level of the individual with existing administrative data to inform their decision-making.

In contrast, for other cases, the boundaries of the policy issue were broader and questions more loosely specified. For example, the enquiry into the consequences of economic downturn and austerity began with the broad aim to utilise linked data to identify changes in frequency and intensity of financial hardship at the local level. Equally, the review of education services began with a general aspiration to better understand changes in the profile of demand. In these cases, formulation of the policy questions and defining and deciding on the scope for data enquiries progressed through a series of incremental, iterative steps. Here, policy-making tended to be in response to emerging policy issues where there were numerous stakeholders advancing competing narratives about the nature of the problems and seeking to shape the range of acceptable policy responses. Thus unlike the healthcare case above, here the framing of the policy questions and legitimate solutions were contested. Despite policy officials' enthusiasm to realise the potential of integrated data, broadly defined questions raised challenges for identifying the types of data that could usefully provide answers. In the education service case, policy officials and service managers initially struggled to conceptualise how the various data on pupils and schools that they held could be exploited. The breadth of policy questions rendered the sources of relevant data that could address the questions as opaque.

Seeking to establish the existence and/or the relevance of data involved conversations with data holders - often managers in departments outside the area

of direct policy interest. This then involved a second-stage of iteration, to establish the validity of the data access request and legitimise the relevance of the data. In the health and wellbeing and the understanding financial hardship case studies, access to data held by other service providers was denied on the grounds that the resource cost of providing data was greater than the perceived benefit to policy. Combining data involves multiple sites where judgements are made about the relevance of data to policy questions that may not be owned or of interest to those that hold the data.

Issues of data relevance are also circumscribed by the divisions of local and national policy responsibilities. In the case of the data enquiry into the impact of economic downturn and austerity, the insights drawn from an analysis of combined datasets on levels of benefit claiming, employment status, county court judgements, household composition, physical health and other factors, showed particular groups of people (in work on low pay) as potentially more exposed to financial hardship. However, the ability of policy officials to action this insight was restricted, as it was deemed outside the scope of local policy. This case illustrated that insights from available and relevant data may not be actionable. This may be for a range of reasons – in this case, local government action was precluded by national government ownership of what transpired to be the issue where action was required.

4.2 Resistance to data use in policy

The case studies provided examples of ways in which the application of data to inform policy was challenged and resisted. For example, policy officials disputed or sought to discredit the legitimacy of data use where they had reservations about its quality. Sometimes claims about poor data quality were substantiated with reference to the purposes for which it had been generated: reservations were expressed around the notion that data collected for one reason should be repurposed for another. On other occasions resistance was focused on the way in which the dataset had been constructed where reservations focused on the validity

of repurposing particular variables. Anticipation of public perceptions about the re-use of data also served to bolster concerns and augment resistance to data use.

In all of the case studies, concern was raised about the potential impact on re-appropriation of the data of missing observations, human error and biases resulting from how they were collected, maintained and stored. In the financial hardship case study, policy officials resisted the inclusion of certain data fields on the grounds that the values they contained may be incorrect. For example, they questioned the quality of some demographic information in one data set where individual characteristics had not been crucial to determining service eligibility. Similarly, in the wellbeing services case, data related to the provision and uptake of these services (e.g. numbers of participants) were perceived to be more systematically collected - and thus more accurate - than evaluation data or data on participants' health outcomes. It was the evaluation and health outcome data however, that was of greater value and significance in the re-appropriation of the data and the potential for linking with other data sets. Thus in both these examples, the extent to which data was considered suitable for reuse was related to the social context in which the data had originally been compiled: the likely motivations underlying the inclusion of particular variables and imputations about the care with which the data set had been constructed.

Further challenges to the validity of data applications for policy were raised in the education services case. Here the legitimacy of repurposing the data was less about the accuracy of the data and more about the validity of extrapolating from it. The example of data on eligibility for free school meals (FSM) illustrates this point. Even where data was perceived to be recorded correctly (i.e. all eligible registrations for FSM were input on data systems), policy officials highlighted that the introduction of universal infant free school meals in 2014 had significantly affected the numbers of parents registering their child's eligibility (Education Policy Institute, 2018). The perceived effect of this policy change was that FSM data had lost its value as an indicator of changed profiles of demand for education services.

In all of the cases, it was not that policy officials lacked curiosity and enthusiasm for harnessing the value of existing data. Indeed, aspirational ideas circulating within and beyond local government (e.g. Shakespeare, 2013; Mayer-Schönberger & Cukier, 2013) about the vast potential of big data, permeated their thinking and motivated their efforts to realise the benefits for policy-making. However, the processes of data curation highlighted that the ability to be curious was tempered by the contexts in which datasets were structured and maintained in local government settings. It was clear in the financial hardship case that a consequence of decisions to hold personal data on clients only for the time that they were service users was that datasets tended to over-represent continuous, and longer term service users thus obscuring patterns in short-term and cyclical service use.

To some extent this was perceived by policy officials to be a consequence of data protection compliance; specifically the requirement to only collect and retain as much personal data as is necessary, and not to reuse data in a way that is incompatible with the original purpose (Information Commissioner's Office, n.d.a; Information Commissioner's Office, n.d.b). Policy officials were reluctant to revisit consent and tended to narrow interpretations of their ability to generate or reuse data. Thus limiting "extensibility" (Mayer-Schönberger & Cukier, 2013, p. :109), whereby the ability of data to have multiple uses is intentionally embedded in data collation protocols. In addition, policy officials were often juggling between two competing narratives about public perceptions of data use by local government. While they recognised a sense of public expectation that they would use available data 'smartly' to innovate and better target services, in practice they were also stifled by anticipation of public reservations about the acceptability of linked data.

As a consequence, emerging awareness of data to answer policy questions did not unproblematically translate into availability of data. Policy makers' sensitivity to data quality and legitimacy, the legality of its use and the anticipated responses of the public could lead to data being rendered inadmissible in integrated data projects. Professional tacit knowledge was used to ground data, counteract its

inaccuracies, navigate its ethical and legal implications and mitigate the likelihood of misreading the insights that it can yield. Data was only admissible where policy professionals could first fill in blanks and inaccuracies with their local knowledge of how things actually are.

4.3 Relationships with data and policy

This final section presents significant aspects of the relationships that effect the intersections between data and policy-making. We first observe that trust is vital to enable integrated data projects to have value in policy settings and then consider how the politics of policy-making impacted data sharing and the terms of engagement for different stakeholders.

Throughout the project collaboration, data was sourced and released in stages as trust in the partnership was built, ethical and legal boundaries established and the value of early analyses realised. For example, in the community health services case, establishing the policy-research relationship led to the policy partners first seeing the potential value of conducting a RER, and then being confident to act on the relationship this showed between patient perception of illness and health management behaviours by collecting attitudinal data that could be linked with secondary health care records.

The data projects proceeded via an abductive approach - flip flopping between patterns emerging in the data and hypotheses, seeking additional insights and testing further hypotheses. For instance, in the example above, having refined the initial scope of the enquiry in the light of the RER, mini hypotheses to explain low service take up by certain patient groups were proposed, tested, discussed and revised in relation to the policy context. Across each of the case studies, the rationale for additional data releases was grounded in the cementing of trust in the partnership and the realisation of benefits from the preceding stages. Thus, the value of the collaborative data enterprise was realised through processes that iteratively established confidence in the partnership.

Sometimes relationships between the project partners were more problematically embedded in the politics of data sharing; for example between levels and departments of government, between different public services, and between the policy partners and the research team. Some data - for example individual-level data on unemployment and take-up of employment services - were held nationally by the Department for Work and Pensions and unavailable to local policy officials on the grounds that it would breach their terms for information governance. Thus relevant data on variance in financial wellbeing was only available to the project in aggregate form.

On one occasion in the community health services project, difficulties in obtaining data from a service provider were attributed to the politics of the relationship between the statutory authority and the provider. Previous experience led the policy partners to suggest that the other party may have been unwilling to share data for fear that the data could be misappropriated and used to monitor their performance. This speaks to the significance of trust and transparency over purpose as well as methods in integrated data projects. Concern about the potential for data to surveil service performance was particularly apparent where ideas about policies – what they intend to achieve, for whom and how – were disputed. For example, in the wellbeing services project, service providers were unwilling to share data with service commissioners where they felt exposed when sharing data showing low volumes of activity without taking into account the quality of provision for vulnerable clients. A further variation on this theme was observed in the review of education services. Here data analysis was sought by service managers where it gave confidence to pursue preferred explanations for changed profiles of demand. Alternative explanations were undermined by questioning data accuracy or by citing particular aspects of policy context.

A final example from the financial hardship case, of the importance of trust was evident in a debate between one of the policy partners and a third sector organisation. The dispute centred on the scale of financial hardship in the local area and the nature of services required in response. Third sector providers made

reference to a range of evidence to support their position. Significantly, the data held by these third sector providers was not made available for integration as they claimed that its collection was conditional on particular sets of expectations for use. Their contention was that the data had been shared with them precisely because they were distinct from local government and a source of support for those wishing to raise grievances about local government. As a result they considered that sharing these data with local authorities would be a breach of trust. This provides a further illustration of how limits on linking data are not restricted to technical issues about the availability or format of data – rather they are shaped by relational considerations around trust and the politics of data and policy-making.

5. DISCUSSION and CONCLUSIONS

The findings and reflections from our project to realise the benefits of data for policy have revealed the ways in which the relevance of data is socially constituted in policy settings; the conditions under which data applications to policy-making can be and are resisted; and the degree to which the relationships between stakeholders at the intersection of data and policy influence what data processes and insights can be considered. Overall, we highlight that variation in the degree to which integrated data and the techniques of data science are able to encroach on policy practice, is contingent on the social context and processes of both data generation and policy-making.

The ambition to utilise the vast quantities of data that local government produces and can access is driven, at least in part, by the motivation to realise the aspirational claims made about big data for policy-making. However, the cases we draw on highlight the first-and-foremost requirement to be problem-oriented in big data applications to policy. Even where we observe the seeming ubiquity of data, there are still circumstances where we have data for which there aren't questions and questions for which we do not have data (Kennedy et al, 2015 boyd

& Crawford, 2012); and it is questions and not data that drive policy calls on evidence.

In contrast to early definitions of big data that focused on the characteristics of data (volume, variety, velocity) with less reference to the purposes to which it could or should be put, we find that where integrated data is applied to policy-making its most defining quality is its ability to be big in value (OECD, 2013; Cowls & Schroeder, 2015). In policy settings the value of data is allied to its ability to provide insight germane to live and pertinent policy and practice preoccupations. We find that the choice of what data to use or collect involves problem-based decisions on what would be indicative of the thing(s) we are trying to understand.

Given this grounding for the potential of data for policy, the social contexts and processes involved in data generation, maintenance and storage become of vital importance. It is these contexts and processes that determine what data can, and what it cannot, represent. We have shown that administrative datasets tend to function primarily as a tool to audit public services; telling us how many services are delivered, for how many people and when. As such, their reuse value is limited where the aim of data applications to policy enquiry is the curious exploration of social phenomena, to understand what could work better, for whom and under what conditions.

The value of integrated data to policy challenges is further exacerbated when consideration is given to the errors and biases data contains as a consequence of how it is arrived at; what priorities are ascribed to its accuracy; and what legitimacy and legality it has when it is repurposed. The implication of these considerations is that the existence of large quantities of data is not an asset in itself to local policy-making. Its value can only be realised if and when the constraints of the social contexts and processes of its production can be mitigated. Even then, we have shown that the potential value of data is conditional on the political context in which policy is being made.

We have shown considerable differences in the contexts in which local policies are made. These contexts are not fixed and static, but highly variable, multifaceted and contingent on the historical trajectory of policy-making in the field. The context shapes ways of acknowledging problems and justifying the solutions to which policy is aimed.

Policy-making takes place on different timescales depending on the mode of policy-making. For instance, whether policy-making is happening as part of a routine programme of on-going review, or in response to an unanticipated shock - such as a public (media) outcry, a change in national or regional policy, or a change in social/economic circumstances - that disrupts routine policy-making processes and 'normal' policy timetables. At any given time, policy concerns can accelerate up through the rankings of priorities, or become suddenly subordinate to other more pressing preoccupations.

Big data analytics, with its focus on quick, novel and exploratory enquiry (Mayer-Schönberger & Cukier, 2013; Hochtl et al, 2016), could be seen to align well with extraordinary and fleet of foot policy-making; often seen as happening at a pace that traditional methods of information generation can't match (Whitty, 2015). However, such an assessment of the potential impact of big data-derived evidence underplays the complexity and politics of policy-making, particularly at points of disruption - for example times of economic downturn and austerity. In our experience, both times of routine policy appraisal and urgent reaction to policy crisis involve, first and foremost, the advancement and debate of ideas about policy, as well as related ideas about data, data science and what constitutes evidence.

The extent to which policy problems and potential options are tightly defined and agreed upon differs in different policy context. Ideas about policy, data and evidence are contained within a political reality that shapes and delimits the boundaries of policy aims; the purpose to which it can be addressed, the extent to which ownership and responsibility over the domain is open or closed, and the

degree of disagreement and dispute among stakeholders over the aims and purpose of policy. The nature of the policy-making context and the issues being explored affects what questions can legitimately be asked of big data and the ways in which the resultant insights are considered as admissible as evidence that can form the basis for decision-making. Issues vary in the degree to which they are contested, how urgent they are, how open, how risky, etc. As a consequence, we find that highly contested local welfare policy has a qualitatively different profile of considerations shaping the 'pull' on data science than, for example, the temporarily more consensual context of local health service provision for patients with a particular chronic condition.

Thus in our exploration of how the practices of data science intersect with the practices and preoccupations of policy, we find a more nuanced and politically contingent call on data science than would be suggested by the rhetoric around data science. Indeed, we suggest that rather than looking at data science as technical aspect of government activity, we should instead see data science as contingent on the realities and political contexts of government practice. Scholarship and practice around these topics must be alert to both the potential impact of data on policy-making but also the ways in which policy-making conditions the potential for data.

REFERENCES

Anderson, C. (2008). The end of theory: Will the data deluge make the scientific method obsolete?. Retrieved December 3, 2015, from http://www.wired.com/2008/06/pb-theory/.

Bergold, J., & Thomas, S. (2012). Participatory research methods: A methodological approach in motion. Forum: Qualitative Social Research, 13(1). doi: 10.17169/fqs-13.1.1801

Boyd, D., & Crawford, K. (2012). Critical questions for big data. Information, Communication & Society, 15(5), 662-679. doi: 10.1080/1369118X.2012.678878 Cairney, P., & Heikkila, T. (2014). A comparison of theories of the policy process. In P. Sabatier & C. Weible (Eds.), Theories of the Policy Process (3rd ed). Chicago: Westview Press.

Carmel, E., & Papadopoulos, T. (2003). 'The new governance of social security in Britain', in Millar, J. (ed.), 'Understanding social security', Bristol: Policy Press, pp. 31-52

Coghlan, D., & Brydon-Miller, M. (Ed.) (2014). The SAGE encyclopedia of action research (Vols. 1-2). London: SAGE Publications Ltd.

Cowls, J., & Schroeder, R. (2015). Causation, correlation, and big data in social science research. Policy and Internet, 7(4), 447-472.

Education Policy Institute. (2018). Evaluation of universal infant free school meals. Retrieved May 7, 2018, from https://epi.org.uk/publications-and-research/evaluation-universal-infant-free-school-meals/.

Giest, S. (2017). Big data for policymaking: Fad or fasttrack? Policy Sciences, 50(3), 367-382, DOI: 10.1007/s11077-017-9293-1

Höchtl, J., Parycek, P., & Schöllhammer, R. (2015) Big data in the policy cycle: Policy decision making in the digital era, Journal of Organizational Computing and Electronic Commerce, 26(1-2), 147-169, DOI: 10.1080/10919392.2015.1125187

Information Commissioner's Office. (n.d.a). Guide to data protection. Retrieved May 7, 2018, from https://ico.org.uk/for-organisations/guide-to-data-protection/.

Information Commissioner's Office. (n.d.b). Guide to the general data protection regulation (GDPR). Retrieved May 7, 2018, from https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/

Janssen, M., Konopnicki, D., Snowdon, J. L., & Adegboyega, A. (2017). Driving public sector innovation using big and open linked data (BOLD). Information Systems Frontiers, 19(2), 189-195.

Janssen, M., & Helbig, N. (2016). Innovating and changing the policy-cycle: Policy-makers be prepared! Government Information Quarterly. Advance online publication. DOI: 10.1016/j.giq.2015.11.009

Jessop, B. (2009). Cultural political economy and critical policy studies, Critical Policy Studies, 3 (3), pp. 336-356.

Kennedy, H., Moss, G., Birchall, C., & Moshonas, S. (2015). Balancing the potential and problems of digital methods through action research: Methodological reflections. Information, Communication & Society, 18(2), 172-186.

Kettl, D. F. (2016). Making data speak: Lessons for using numbers for solving public policy puzzles. Governance, 29(4), 573-579.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, 1(1), 1–12. DOI: 10.1177/2053951714528481

Malomo, F., & Sena, V. (2016). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. Policy and Internet, 9(1), 7-27.

Markus, M. L., & Topi, H. (2015). Workshop on the social economic and workforce implications of big data analytics and decision making: Report on a research agenda setting workshop. Retrieved March 15, 2016, from https://www.bentley.edu/files/2015/10/08/BigDataWorkshopFinalReport.pdf.

Mayer-Schönberger, V., & Cukier, K. (2013). Big data. A revolution that will transform how we live, work and think. London: John Murray.

Organisation for Economic Co-operation and Development. (2013). Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "big data". Retrieved 7 May, 2018, from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP(2012)9/FINAL&docLanguage=En.

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. Review of Policy Research, 31(4), 343-348.

Schroeder, R. (2014). Big Data: towards a more scientific social science and humanities? In M. Graham & W. H. Dutton (Eds.), Society and the Internet (pp. 164-176). Oxford: Oxford University Press.

Shakespeare, S. (2013). Shakespeare Review: An independent review of public sector information. Retrieved March 15, 2016, from https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information.

Stone, D. A. (1989). Causal stories and the formation of policy agendas. Political Science Quarterly, 104(2), 281-300.

Whitty, C. (2015). What makes an academic paper useful for health policy? BMC Medicine, 13(301), 1-5. doi: 10.1186/s12916-015-0544-8

Parliament.UK. (2015). Making big data available: Key issues for the 2015 Parliament. Retrieved May 22, 2018, from http://www.parliament.uk/business/publications/research/key-issues-parliament-2015/technology/big-data/

All was well.